Ederson Schmeing

Seleção dinâmica de classificadores baseada em medidas de complexidade

Cascavel-PR

2022

Ederson Schmeing

Seleção dinâmica de classificadores baseada em medidas de complexidade

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-Graduação em Ciência da Computação (PPGComp) da Universidade Estadual do Oeste do Paraná – Unioeste, campus de Cascavel.

Universidade Estadual do Oeste do Paraná – Unioeste – Cascavel

Centro de Ciências Exatas e Tecnológicas – CCET

Programa de Pós-Graduação em Ciência da Computação – PPGComp

Orientador: Dr. André Luiz Brun

Coorientador: Dr. Ronan Assumpção Silva

Cascavel-PR

2022

Ederson Schmeing
Seleção dinâmica de classificadores baseada em medidas de complexidade/ Ederson Schmeing. – Cascavel-PR, 2022-
79p. : il. (algumas color.) ; 30 cm.

Orientador: Dr. André Luiz Brun

Dissertação (Mestrado)– Universidade Estadual do Oeste do Paraná – Unioeste – Cascavel
Centro de Ciências Exatas e Tecnológicas – CCET
Programa de Pós-Graduação em Ciência da Computação – PPGComp, 2022.

1. Sistemas de Múltiplos Classificadores 2. Seleção Dinâmica de Classificadores. 3. Dificuldade do Problema de Classificação. 4. Medidas de Complexidade. I. André Luiz Brun. II. Universidade Estadual do Oeste do Paraná – Unioeste – Cascavel. III. Programa de Pós-Graduação em Ciência da Computação - PPGComp. IV. Seleção dinâmica de classificadores baseada em medidas de complexidade

Ederson Schmeing

Seleção dinâmica de classificadores baseada em medidas de complexidade

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-Graduação em Ciência da Computação (PPGComp) da Universidade Estadual do Oeste do Paraná – Unioeste, campus de Cascavel.

Trabalho aprovado. Cascavel-PR, 26 de agosto de 2022:

André L. Brun

Dr. André Luiz Brun
Orientador

Ronan Assumpção Silva

Dr. Ronan Assumpção Silva
Coorientador
Instituto Federal do Paraná (IFPR)

Adair Sta. Catarina

Dr. Adair Santa Catarina
Universidade Estadual do Oeste do Paraná (UNIOESTE) - Campus de Cascavel

Rodolfo Botto de Barros Garcia

Dr. Rodolfo Botto de Barros Garcia
Universidade Federal de Sergipe (UFS)

Cascavel-PR
2022

Este trabalho é dedicado à minha batalhadora mãe Cleide Aparecida Wons que sempre lutou incansavelmente para criar a mim e meu irmão com recursos disponíveis e da melhor forma possível.

Agradecimentos

Agradeço primeiramente a Deus, pela vida e por me dar saúde, permitindo assim a realização de projetos como este. Quero agradecer imensamente ao professor André Luiz Brun por todo apoio e tempo despendido ao longo desta jornada, com ensinamentos, orientações e conselhos. Sou muito grato a ele pela confiança na realização deste trabalho, pela paciência e motivação, sobretudo nos momentos difíceis. Ao professor Ronan Assumpção Silva que, na fase final dos trabalhos, foi extremamente atuante para encerramento das atividades da pesquisa. Aos demais membros da banca, professor Rodolfo Botto de Barros Garcia e Adair Santa Catarina, pelas contribuições que enriqueceram o trabalho, meu muito obrigado.

Meus agradecimentos a todos os envolvidos no programa, ao professor Luiz Antonio Rodrigues, por conduzi-lo com maestria. A todos os professores, principalmente os que de alguma forma contribuíram para a realização deste projeto: Marcio Seiji Oyamada, Victor Francisco Araya Santander, Rosangela Villwock, Guilherme Galante e Adair Santa Catarina.

Deixo um enorme agradecimento ao professor Ivonei Freitas da Silva, ademais das disciplinas ministradas, me apresentou aos professores do Programa, que me proporcionou o conhecimento do meu futuro orientador André, que acreditou e confiou na minha vontade de ingressar no Programa e capacidade de cumprir as atividades do mesmo. Também deixo meu agradecimento ao professor Wesley Klewerton Guêz Assumção que, além das disciplinas lecionadas, logo no início do caminho, nos introduziu profundamente no meio cientifico, nos orientando e conduzindo a produzir e publicar artigos científicos.

Ter verdadeiros mestres ao seu lado é uma das maiores conquistas para se chegar longe. Por isso, não posso deixar de mencionar, neste espaço, os professores Aníbal Mantovani Diniz e Evandro Carlos Pessini, que, com sua grande competência, compartilharam comigo seu vasto conhecimento nas aulas da graduação e da especialização. Ademais, o professor Evando me deu a honra de ser meu orientador em ambos os cursos, o que levarei, para sempre, como uma oportunidade única em minha vida acadêmica. Agradeço, ainda, ao meu superior no Banco do Brasil, o gerente geral Valmir Matiello, por seu apoio. Graças a ele e aos professores citados, obtive a recomendação ao programa - o incentivo de extrema importância para o início desta jornada.

Minha gratidão também vai, com muito carinho, para Ana Julia Lage Muniz de Souza dos Anjos pelos comentários e sugestões que auxiliaram na revisão do trabalho e que, certamente, qualificam e conferem melhorias em nosso texto.

Não seria possível a concretização desta pesquisa sem o apoio incondicional da

minha esposa Isabela Maria Chalegre dos Anjos Schmeing, que me ajudou muito a trilhar este caminho, incentivando, motivando e entendendo algumas privações e ausências. Agradecimento também à toda minha família pelo carinho e incentivo. Ao meu irmão Anderson Luis Schmeing e sua esposa que estão cuidando da nossa mãe com grande zelo.

Agradeço aos colegas do PPGComp: Daniele Wolfart, Diogo Carlos Paganini Domingos, Diogo do Nascimento Paza, Elias Abreu Domingos da Silva, Guilherme Luciano Donin Villaca e Marcio Veronez, pelo aprendizado e desenvolvimento mútuo. Assim como, todos meus amigos e colegas que, de alguma forma, se envolveram e contribuíram para o cumprimento desse desafio.

Por fim, agradeço a instituição Banco do Brasil pela bolsa concedida, que permitiu abonar as horas faltantes na jornada de trabalho devido as atividades acadêmicas.

"O homem é feito de tal modo que quando alguma coisa incendeia a sua alma, as impossibilidades desaparecem.
(Jean de La Fontaine)

Resumo

SCHMEING, Ederson. **Seleção dinâmica de classificadores baseada em medidas de complexidade**. Orientador: Dr. André Luiz Brun. 2022. 79f. Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual do Oeste do Paraná, Cascavel – Paraná, 2022.

A classificação de objetos é uma tarefa da área de Reconhecimento de Padrões. Ela consiste em atribuir uma classe, dentre várias possíveis, a um objeto desconhecido. Esse processo é realizado por meio de classificadores. Eles são funções matemáticas formuladas a partir do treinamento do(s) classificador(es), que fazem uso dos dados de um determinado problema para realizar a aprendizagem. Classificadores monolíticos encontram dificuldade em cenários mais difíceis de classificação. Eles podem não atender a todo o espaço de variabilidade do problema. Um Sistema de Múltiplos Classificadores (SMC) têm sido uma opção para tentar melhorar desempenho da tarefa de classificação em relação à abordagem monolítica. Para isso, o SMC emprega diversos classificadores especialistas em regiões distintas do problema. A ideia é que, a partir de um conjunto de diferentes classificadores se consiga absorver as variabilidades apresentadas pelos problemas. No entanto, nem sempre todos eles são empregados na etapa de classificação, sendo necessária a etapa de seleção de classificadores do conjunto. Para esta tarefa a seleção dinâmica dos classificadores tem se mostrado promissora. A arquitetura com que são construídos os sistemas de classificação influencia no seu desempenho. Outra dificuldade do problema de classificação está relacionada à complexidade dos dados. Ela impacta na forma que os classificadores atuam e as informações acerca dela é obtida a partir das medidas de complexidades. Estudos demostram que para resolver um problema difícil de classificação, uma melhor compreensão da complexidade dos dados pode ser interessante para seleção de classificadores. Assim, neste trabalho, avaliamos as medidas de complexidade como critérios base para a seleção dinâmica de classificadores. A modificação das abordagens clássicas da literatura resulta em maior acurácia de $39,51\%$ dos casos quando adicionamos as medidas de complexidade. Enquanto a abordagem original corresponde a $28,4\%$. Empates estiveram presentes em $35,8\%$ dos casos testados. Este trabalho também apresenta o ganho de informação de 14 medidas de complexidade, o que sugere algumas desvantagens do uso de algumas delas.

Palavras-chave: Sistemas de Múltiplos Classificadores, Seleção Dinâmica de Classificadores, Dificuldade do Problema de Classificação, Medidas de Complexidade

Abstract

SCHMEING, Ederson. **Dynamic selection of classifiers based on complexity measures**. Orientador: Dr. André Luiz Brun. 2022. 79f. Dissertação (Mestrado em Ciência da Computação) – Universidade Estadual do Oeste do Paraná, Cascavel – Paraná, 2022.

Object classification is a task in the field of pattern recognition. It consists of assigning a class, among several possible ones, to an unknown object. This process is accomplished by classifiers. They are mathematical functions built from the training of the classifiers, which make use of the data of a given problem to obtain learning. Monolithic classifiers find difficulty in more difficult classification scenarios. They may not cover the full range of variability of the problem. Multiple Classifier Systems (MCS) have been an option in order to improve the performance of the classification task before the monolithic approach. For this, the MCS's employ several specialist classifiers in different regions of the problem. The idea is that, from a set of different classifiers, it is possible to absorb the variability presented by the problems. However, not all of them are always used in the classification step, being necessary the step of selection of classifiers from the set. For this task, the dynamic selection has shown to be promising. The architecture with which classification systems are built influences their performance. The difficulty of the classification problem is also related to the complexity of the data. It impacts the way the classifiers act and information about it is obtained from the complexity measures. Studies show that to solve a difficult classification problem, a better understanding of data complexity can be interesting for classifier selection. Thus, in this work, we evaluate the complexity measures as selection criteria for the dynamic selection of classifiers. Modifying classical literature approaches results in higher accuracy of 39.51% of the cases by adding the complexity measures, while maintaining the original approach corresponds to 28.4%. Ties were present in 35.8%. This work also presents the information gain of 14 complexity measures, which suggest some drawbacks of using some of them.

Keywords: Multiple Classifier Systems, Dynamic Classifiers Selection, Classification Problem Difficulty, Complexity Measures

Lista de ilustrações

Figura 1 – Etapas de um SMC. Adaptado de (BRITTO JR.; SABOURIN; OLIVEIRA, 2014; CRUZ; SABOURIN; CAVALCANTI, 2018) 23
Figura 2 – Ilustração do método *Bagging* para geração de subconjuntos 25
Figura 3 – Ilustração do método *Boosting* para geração de subconjuntos 26
Figura 4 – Ilustração do método RSS para geração de subconjuntos 27
Figura 5 – Seleção estática de classificadores. Adaptado de (CRUZ; SABOURIN; CAVALCANTI, 2018) 29
Figura 6 – Seleção dinâmica de um classificador. Adaptado de (CRUZ; SABOURIN; CAVALCANTI, 2018) 30
Figura 7 – Seleção dinâmica de um conjunto de classificadores. Adaptado de (CRUZ; SABOURIN; CAVALCANTI, 2018) 30
Figura 8 – Exemplo prático do método OLA (*Overall Local Accuracy*) 32
Figura 9 – Procedimento do método LCA (*Local Class Accuracy*) 33
Figura 10 – Ilustração do funcionamento do método *KNORA-Eliminate (KN-E)* . . 35
Figura 11 – Ilustração do funcionamento do método *KNORA-Union (KN-U)* 35
Figura 12 – Categorias de saídas dos classificadores em um SMC 38
Figura 13 – Nível de confiança dos classificadores C_1, C_2 e C_3 de cada uma das w classes para amostra xt_i. 39
Figura 14 – Voto majoritário adaptado de (KUNCHEVA, 2014) 40
Figura 15 – Ilustração do funcionamento do voto majoritário ponderado 41
Figura 16 – Medida de Complexidade $F2$ 47
Figura 17 – Cenário em que um classificador linear erra ao classificar as três amostras 50
Figura 18 – Árvore geradora mínima construída com base no conjunto de dados de entrada 52
Figura 19 – Medida de complexidade $N2$ 53
Figura 20 – Visão geral da estratégia 63
Figura 21 – Espaço de características e representação do espaço de complexidade . 64
Figura 22 – Representação do novo espaço de característica, combinando a característica original com as medidas de complexidade. 68
Figura 23 – Gráfico comparativo entre os desempenhos das abordagens com e sem a adição das medidas de complexidade. Em verde as vitórias da técnica que adota as medidas, em azul as derrotas e em laranja os empates . . 70

Lista de tabelas

Tabela 1 – Combinação do nível de confiança dos classificadores 40
Tabela 2 – Funcionamento da regra *Borda Count* 42
Tabela 3 – Principais características das bases usadas nos experimentos 62
Tabela 4 – Métodos de Seleção Dinâmica (DS) incluídos na análise comparativa . 65
Tabela 5 – Classificação média de ganho de informação para as medidas de complexidade 67
Tabela 6 – Comparação das performances médias nas 20 execuções. Os desvios padrão correspondentes são dados entre parênteses. 71

Lista de abreviaturas e siglas

SMC	Sistema de Múltiplos Classificadores (MCS - Multiple Classifier System)
EoC	Ensemble of Classifier
DS	Dynamic Selection
LCA	Local Classifier Accuracy
OLA	Overall Local Accuracy
KNORA	K-Nearests Oracles
KNORA-E	K-Nearests Oracles-Eliminate
KNORA-U	K-Nearests Oracles-Union
SDES	Sorting-Based Dynamic Classifier Ensemble Selection
NN	Nearest Neighbors
KNN	K-Nearest Neighbors
MST	Minimum Spanning Tree
DCol	Data Complexity Library

Lista de símbolos

X	Conjunto de dados
S	Subconjuntos de dados
X_{train}	Conjunto de dados de treinamento
X_{val}	Conjunto de dados de validação
X_{test}	Conjunto de dados de teste
n	Número de instâncias do conjunto de dados
d	Número de dimensões (atributos) do conjunto de dados
C	*Pool* de classificadores
C'	Subconjunto de classificadores
C_i	i-ésimo classificador
M	Número de classificadores presentes no *pool* C
M'	Número de classificadores presente no subconjunto de classificadores C'
x	Vetor de atributos de uma instância do conjunto de dados
x_i	i-ésima instância de um conjunto de dados ($x_i \in X$)
xv_i	i-ésima instância do conjunto de validação ($xv_i \in X_{val}$)
xt_i	i-ésima instância de teste cuja classe é desconhecida ($xt_i \in X_{test}$)
$\tilde{xt}_i$	Saída de uma instância de teste
f_j	j-ésimo atributo de uma instância
x_{ij}	Valor do j-ésimo atributo da i-ésima instância.
w	Vetor de classes
w_i	i-ésima classe do vetor w
K	Número de vizinhos na região de competência
Viz	Vizinhança de tamanho K de uma instância qualquer
Sig	Assinatura de complexidade de um conjunto

Sumário

1 INTRODUÇÃO 17
1.1 Hipótese, questão de pesquisa e motivação 19
1.2 Proposta 20
1.3 Objetivo 20
1.4 Contribuições 20
1.5 Estrutura do trabalho 20

2 CLASSIFICAÇÃO 22
2.1 Geração de classificadores 24
2.1.1 *Bagging* 24
2.1.2 *Boosting* 24
2.1.3 *Random subspaces* 25
2.1.4 Geração direcionada por complexidade 26
2.1.5 Diversidade entre classificadores 28
2.2 Seleção de classificadores 28
2.2.1 Seleção estática 29
2.2.2 Seleção dinâmica 30
2.2.2.1 Oráculo 31
2.2.2.2 Seleção dinâmica do classificador 31
2.2.2.3 Seleção baseada em *ranking* 32
2.2.2.4 Acurácia local total - OLA 32
2.2.2.5 Acurácia local da classe - LCA 32
2.2.2.6 Seleção a priori 33
2.2.2.7 Seleção a posteriori 33
2.2.2.8 Seleção dinâmica de conjunto de classificadores 34
2.2.2.9 K oráculos mais próximos - KNORA 34
2.2.2.10 Seleção baseada em diversidade e acurácia 35
2.2.2.11 Seleção baseada em diversidade - SDES 36
2.3 Combinação de classificadores 37
2.3.1 Produto 38
2.3.2 Soma 38
2.3.3 Máximo 38
2.3.4 Mínimo 39
2.3.5 Média 39
2.3.6 Mediana 39

2.3.7 Cenário de aplicação das regras . . . 39
2.3.8 Voto majoritário . . . 40
2.3.9 Voto majoritário ponderado . . . 40
2.3.10 *Borda count* . . . 41
2.3.11 *Ranking* . . . 42
2.4 Considerações finais . . . 42

3 MEDIDAS DE COMPLEXIDADE . . . 44
3.1 Medidas baseadas em atributos . . . 44
3.1.1 Relação máxima do discriminante de fischer ($F1$) . . . 44
3.1.2 Relação máxima do discriminante de fischer do vetor direcional ($F1v$) . . . 46
3.1.3 Volume da região de sobreposição ($F2$) . . . 46
3.1.4 Eficiência máxima por atributo individual ($F3$) . . . 47
3.1.5 Eficiência coletiva dos atributos ($F4$) . . . 48
3.2 Medidas de linearidade . . . 49
3.2.1 Soma da distância de erro por programação linear ($L1$) . . . 49
3.2.2 Taxa de erro do classificador linear ($L2$) . . . 50
3.2.3 Não linearidade de um classificador linear ($L3$) . . . 51
3.3 Medidas de vizinhança . . . 51
3.3.1 A fração de pontos na região de fronteira ($N1$) . . . 51
3.3.2 Proporção da distância intra/extra da classe do vizinho mais próximo ($N2$) 52
3.3.3 Taxa de erro do classificador do vizinho mais próximo ($N3$) . . . 53
3.3.4 Não linearidade do classificador do vizinho mais próximo ($N4$) . . . 54
3.3.5 Fração de esferas de cobertura máxima do dados ($T1$) . . . 54
3.4 Medidas de dimensionalidade . . . 54
3.4.1 Número médio de pontos por dimensão ($T2$) . . . 55
3.4.2 Número médio de dimensões PCA por pontos ($T3$) . . . 55
3.4.3 Relação entre a dimensão PCA e a dimensão original ($T4$) . . . 55
3.5 Medidas de balanço entre classes . . . 56
3.5.1 Entropia da proporções das classe ($C1$ - $B1$) . . . 56
3.5.2 Relação de balanço multi-classes ($C2$ - $B2$) . . . 56
3.6 Medidas de correlação de atributos . . . 57
3.6.1 Correlação de atributos perante a saída ($C1$) . . . 57
3.6.2 Correlação média de atributos perante a saída ($C2$) . . . 58
3.6.3 Eficiência de atributo individual ($C3$) . . . 58
3.6.4 Eficiência de atributos coletivos ($C4$) . . . 59
3.7 Demais medidas . . . 59
3.7.1 Densidade ($D1$) . . . 59
3.7.2 Volume da vizinhança local ($D2$) . . . 59
3.7.3 Densidade da classe na região de sobreposição ($D3$) . . . 60

3.8 Considerações finais 60

4 METODOLOGIA 61
4.1 Conjunto de dados 61
4.2 Geração do conjunto de classificadores base (A) 62
4.3 Opinião do conjunto de classificadores base para validação (B) . . . 63
4.4 Assinatura de complexidade de instâncias de validação (C) 64
4.5 Assinatura de complexidade de instâncias de teste (D) 64
4.6 Seleção de classificadores (E) 65
4.7 Avaliando as medidas de complexidade 66
4.8 Considerações Finais 68

5 EXPERIMENTO E RESULTADOS 69
5.1 Considerações finais 72

6 CONCLUSÕES 73

REFERÊNCIAS 75

1 Introdução

O Reconhecimento de Padrões é a área de conhecimento que tem por objetivo a classificação ou o reconhecimento de objetos em categorias ou classes considerando conhecimentos prévios (DUDA; HART; STORK, 2001; KUNCHEVA, 2014). Esse conhecimento é chamado de aprendizado, pois o classificador armazena informações acerca das classes e as representa através de um modelo.

Uma classe contém objetos semelhantes que são descritos por características, também chamadas de atributos. Um conjunto de dados não homogêneo é formado por objetos de diversas classes (KUNCHEVA, 2014; LORENA et al., 2021). Os classificadores são os elementos responsáveis por analisar um objeto desconhecido através da similaridade de suas características com àquelas conhecidas previamente e predizer a classe do novo exemplar.

O resultado desejado no processo de classificação é que os classificadores possam fazer a atribuição da classe com precisão. Para tal, a escolha de um classificador apropriado para o problema é essencial. Todavia, devido à variabilidade dos problemas de classificação, um classificador escolhido para uma determinada instância pode não obter bons resultados em outras no espaço do problema.

Segundo Wolpert e Macready (1997), não existe uma única técnica de aprendizado de máquina adequada para todos os tipos de problemas. Esse teorema é chamado de *"No Free Lunch Theorem"*. Portanto, uma vez que não existe um classificador ideal, uma possibilidade é agrupar diversos classificadores como alternativa para tarefas de classificação. Essa é a premissa da combinação de classificadores também apresentada na literatura como Sistemas de Múltiplos Classificadores (SMC) ou *Ensemble Learning*.

A combinação de classificadores é uma técnica de aprendizagem que integra classificadores com o propósito de melhorar o desempenho e a precisão geral, de forma a superar o resultado alcançado por um único classificador, apoiando-se no argumento de que as várias opiniões de um grupo são melhores do que a opinião de um único especialista.

Um SMC é composto por três etapas: geração, seleção e integração. Na primeira, um *pool* de classificadores é treinado para criar um conjunto preciso e diversificado. Na segunda etapa, é selecionado o classificador ou um conjunto de classificadores considerados mais competentes. Na última etapa, as opiniões dos especialistas selecionados são combinadas para se obter a decisão final do sistema de acordo com alguma regra de combinação.

A geração dos classificadores pode ser realizada de forma homogênea ou heterogênea na construção do *pool*. A primeira forma utiliza algoritmos de um mesmo tipo, treinados

em diferentes conjuntos de dados. Na segunda estratégia, entretanto, os algoritmos de aprendizagem são diferentes, mas treinados sobre o mesmo conjunto de dados.

A etapa de seleção pode ser realizada de forma estática ou dinâmica. Nas técnicas de seleção estática, um ou mais classificadores são selecionados durante a fase de treinamento, de acordo com um critério de seleção estimado no conjunto de dados de validação. O mesmo conjunto de classificadores é usado para prever a classe de todas as amostras de teste.

Nas técnicas de seleção dinâmica, um único classificador ou um *ensemble* é selecionado especificamente para classificar cada nova instância. Com base em um conjunto de classificadores, as técnicas de seleção dinâmica consistem em encontrar um único classificador ou um conjunto de classificadores para uma amostra de teste específica. A premissa das técnicas de seleção dinâmica é que cada classificador é um especialista em regiões distintas do espaço do problema (KO; SABOURIN; BRITTO JR., 2008; BRITTO JR.; SABOURIN; OLIVEIRA, 2014; CRUZ; SABOURIN; CAVALCANTI, 2018).

Na etapa de combinação ou integração, uma decisão é tomada com base nas saídas obtidas pelos classificadores selecionados conforme alguma regra de combinação (KITTLER, 1998; GUNES et al., 2003; KUNCHEVA, 2014).

A eficiência e o desempenho de um sistema de reconhecimento de padrões são diretamente influenciados pela arquitetura adotada. Porém, outro fator muito importante, o qual também impacta no desempenho desses sistemas, são os dados a serem analisados.

Estudos mostram que os dados influenciam no comportamento do modelo de aprendizagem e que a complexidade do problema a ser resolvido é uma medida que interfere diretamente na qualidade, na eficiência e na acurácia dos classificadores (HO; BASU, 2002; MACIà; ORRIOLS-PUIG; BERNADó-MANSILLA, 2010).

Um conjunto em que os dados estejam mais bem distribuídos no espaço de característica tende a ser mais simples de se resolver, principalmente se linearmente separável. Por outro lado, em conjuntos de dados intrincados, nos quais há sobreposição de classes, a tendência é estes serem mais complexos.

Uma forma de tentar estimar o grau de dificuldade de um problema são as medidas de complexidades. Tais descritores buscam avaliar o quão discriminativos são os atributos; verificar até que ponto duas classes são separáveis linearmente; encontrar a região de fronteira de decisão entre as classes; indicar a dispersão dos dados; mensurar o desbalanceamento entre as classes; considerar a correlação dos valores dos atributos com as saídas; e determinar a densidade das classes (HO; BASU, 2002; SOTOCA; SáNCHEZ; MOLLINEDA, 2005; LORENA et al., 2018; LORENA et al., 2019).

Para isso, os métodos de seleção dinâmica presentes na literatura procuram medir a competência dos classificadores, buscando selecionar um ou mais elementos que sejam

apropriados para classificar cada novo padrão. A competência é geralmente calculada levando em consideração uma região local em torno do novo padrão para encontrar seus vizinhos. Para tanto, diferentes critérios de seleção são utilizados. Há métodos que usam *ranking* (SABOURIN et al., 1993), acurácia (WOODS; KEGELMEYER; BOWYER, 1997; GIACINTO; ROLI, 1999; GIACINTO; ROLI, 2001; DIDACI et al., 2005), classificação aleatória (WOLOSZYNSKI et al., 2012), ordenação (YAN et al., 2013), probabilidade (WOLOSZYNSKI; KURZYNSKI, 2011), relacionamento (GIACINTO; ROLI, 2001), meta-aprendizagem (CRUZ et al., 2015; CRUZ; SABOURIN; CAVALCANTI, 2017), oráculo (KO; SABOURIN; BRITTO JR., 2008), complexidade (BRUN et al., 2016; BRUN et al., 2018), diversidade (SANTANA et al., 2006; KUNCHEVA; RODRIGUEZ, 2007), ambiguidade (SANTOS; SABOURIN; MAUPIN, 2007) e manipulação dos dados (XIAO; HE, 2009) - tudo a fim de determinar os classificadores selecionados.

A complexidade dos dados como critério de seleção encontra-se em estágio inicial, carecendo de pesquisas que analisem se as informações fornecidas pelas medidas de complexidade têm relação com o processo de seleção de classificadores. No estudo de Brun et al. (2016) notou-se que considerar a similaridade em termos de complexidade da vizinhança das instâncias é uma estratégia promissora.

Segundo Ho e Basu (2002) e Macià, Orriols-Puig e Bernadó-Mansilla (2010), para se resolver um problema de classificação, compreender melhor a complexidade dos conjuntos de dados em que os classificadores são treinados e da região de competência do classificador é uma estratégia interessante para a seleção de classificadores. Assim sendo, nesta pesquisa buscou-se avaliar se as medidas de complexidade poderiam contribuir como critério de seleção para seleção dinâmica de classificadores em um SMC.

1.1 Hipótese, questão de pesquisa e motivação

A hipótese deste trabalho consiste em considerar que as medidas de complexidade pode contribuir para a seleção dinâmica de classificadores, obtendo melhores resultados diante de estratégias clássicas presentes na literatura.

A principal questão da pesquisa é: O uso da complexidade dos dados como critério para seleção do classificador ou dos classificadores de forma dinâmica em um sistema de múltiplos classificadores pode contribuir para tarefas de classificação? A motivação da pesquisa baseia-se na premissa de que classificadores treinados em conjuntos de dados com complexidades distintas tendem a ter desempenhos diferentes de acordo com a complexidade do problema. Dessa forma, ao se identificar um classificador treinado em uma região com complexidade similar à da vizinhança da instância de teste, espera-se que este esteja melhor preparado para classificar o exemplar em questão.

Além disso espera-se que, ao explorarmos a adoção de descritores de complexidade

em cenários de seleção dinâmica de classificadores possamos lançar alguma luz sobre a análise dessa relação, bem como contribuir com a área de reconhecimento de padrões.

1.2 Proposta

Desenvolvimento e avaliação de uma estratégia para seleção dinâmica de conjuntos de classificadores com base nas informações sobre a complexidade dos dados usados no treinamento dos classificadores e a acurácia dos mesmos na região de vizinhança do padrão de teste. Para isso, foram analisadas as vizinhanças das instâncias pertencentes aos subconjuntos de teste no espaço de complexidade, ou seja, a nova amostra era projetada dentro do conjunto de validação. Em seguida avaliou-se a aplicabilidade da proposta construída, comparando seu desempenho, em termos de acurácia, perante métodos presentes na literatura.

1.3 Objetivo

Este trabalho teve como objetivo avaliar a contribuição das métricas de complexidade como critério de seleção para seleção dinâmica classificadores.

1.4 Contribuições

- As métricas de complexidade que contribuem para a técnica DS;
- Métricas de complexidade que não contribuem para a técnica DS;
- Uma análise experimental de 21 conjuntos de dados contendo apenas duas classes possíveis, na qual avaliamos o desempenho da técnica proposta frente a quatro técnicas clássicas de DS.

1.5 Estrutura do trabalho

Depois do Capítulo 1 é apresentada a Revisão da Literatura, a qual está dividida em outros dois capítulos. No capítulo 2 são apresentados conceitos sobre a classificação de objetos e o Sistema de Múltiplos Classificadores (SMC). No capítulo 3 são apresentadas e detalhadas várias medidas de complexidade dos dados que têm como objetivo mostrar um entendimento maior sobre o relacionamento dos dados.

A metodologia é detalhada no Capítulo 4, em que é descrito o funcionamento do método proposto de forma genérica. São apresentadas cada uma de suas etapas, desde a divisão do conjunto de dados, passando pela geração do *pool* até a seleção dos classificadores.

Já o capítulo 5 apresenta os experimentos e os resultados. E, por fim, o capítulo 6 detalha a conclusão dos resultados. O apêndice **??** apresenta a primeira versão da metodologia.

2 Classificação

A classificação consiste em atribuir uma classe apropriada a um objeto com base em uma abstração gerada usando um conjunto de padrões de treinamento. É o processo de identificar a qual classe um objeto desconhecido pertence. A classificação é uma das principais tarefas da área de Reconhecimento de Padrões.

O papel de um classificador é atribuir classes aos objetos ainda não classificados. Ele pode ser representado como um conjunto de funções matemáticas, aproximadas, encontradas através do treinamento com dados que permitem fazer previsões. Esse processo é realizado pela aprendizagem de máquina (DUDA; HART; STORK, 2001; KUNCHEVA, 2014).

Um sistema de reconhecimento composto por um único classificador é chamado de monolítico ou individual. Ele é responsável por classificar todos os objetos desconhecidos. Porém, atribuir a esse classificador a tarefa de cobrir toda a variabilidade do problema, pode o desempenho do reconhecimento ser prejudicado, pois este classificador dificilmente será capaz de cobrir toda variabilidade presente em um dado problema. Além disso, escolher o melhor classificador para solucionar todo o espaço do problema é uma tarefa desafiadora.

A expectativa é que o classificador faça a atribuição da classe com precisão, com base nas características do objeto. Para isso, a escolha de um classificador adequado é fundamental para a classificação de um conjunto de instâncias. Entretanto, o classificador escolhido para uma tarefa pode não ter o mesmo desempenho para outra. Identificar o melhor classificador, de acordo com os dados a serem classificados, é um dos grandes desafios de classificação (KO; SABOURIN; BRITTO JR., 2008; KUNCHEVA, 2014).

Uma alternativa aos classificadores monolíticos é unir esforços de vários classificadores para tentar mitigar o problema da escolha do classificador por tarefa, de forma a cobrir toda ou a maior parte da variabilidade do espaço de classificação, minimizar os erros e aumentar o desempenho e a precisão desses sistemas. Essa é a ideia dos Sistemas de Múltiplos Classificadores (SMC), também conhecidos como *Ensemble Learning*, que consiste em uma técnica de aprendizagem que combina vários classificadores fracos, a fim de produzir um classificador forte, diminuindo o viés e a variância. Essa técnica é eficaz e poderosa e permite aumentar consideravelmente o nível de precisão da classificação em comparação à utilização de sistemas monolíticos (KITTLER, 1998; GUNES et al., 2003; KUNCHEVA, 2014; BRITTO JR.; SABOURIN; OLIVEIRA, 2014; CRUZ; SABOURIN; CAVALCANTI, 2018).

O sucesso de se utilizar um SMC depende da diversidade do conjunto de classificadores. Para tanto, os integrantes do conjunto devem ser diferentes, passíveis de cometer

erros diversos não relacionados e complementares, uma vez que não há razão para se combinar especialistas que sempre apresentem a mesma opinião (SHIPP; KUNCHEVA, 2002; KUNCHEVA; WHITAKER, 2003; BROWN et al., 2005). Um sistema de múltiplos classificadores é composto por três etapas: 1) geração; 2) seleção; e 3) integração, que são representadas na Figura 1 (BRITTO JR.; SABOURIN; OLIVEIRA, 2014; CRUZ; SABOURIN; CAVALCANTI, 2018).

Na Figura 1, o SMC é representado por suas três etapas. Inicialmente, por meio de um conjunto de dados de treinamento X_{train} é gerado um *pool* de M classificadores $C = \{C_1, C_2, ..., C_M\}$, os quais são, então, levados a um segundo passo para a seleção daqueles M' classificadores que se mostrarem mais promissores, usando um conjunto de validação X_{val}, onde $(M' \leq M)$. Por último, as saídas dos classificadores são combinadas para se obter uma decisão final.

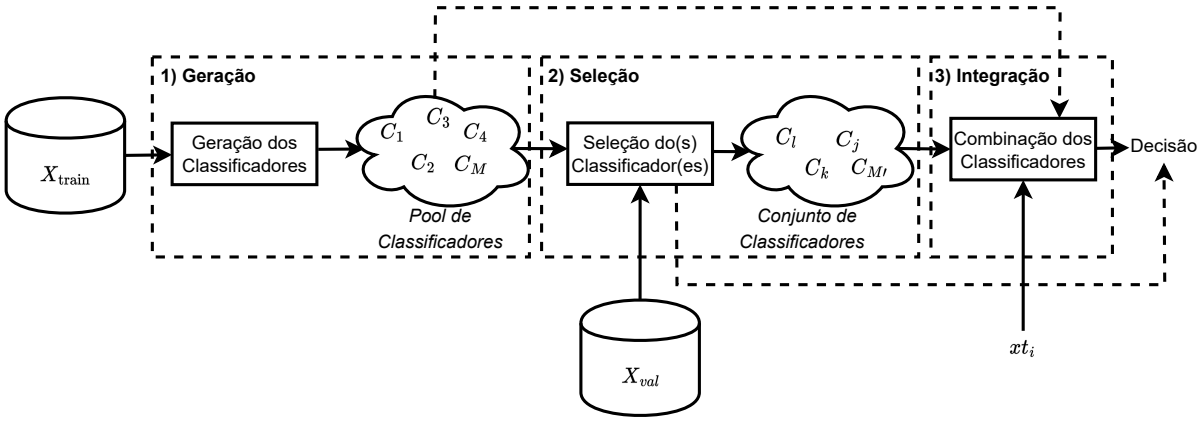


Figura 1 – Etapas de um SMC. Adaptado de (BRITTO JR.; SABOURIN; OLIVEIRA, 2014; CRUZ; SABOURIN; CAVALCANTI, 2018)

A primeira etapa (Figura 1-1) é responsável pela geração de um *pool* de classificadores com base em alguma estratégia de forma a construir um conjunto de classificadores precisos e diversos. Tal processo é descrito com mais detalhes na Seção 2.1.

Na etapa de seleção (Figura 1-2), busca-se escolher o classificador mais competente ou um ensemble de classificadores (*Ensemble of Classifiers* - EoC) mais competentes para o problema. Caso a opinião de todos os classificadores do conjunto seja levada em conta, a etapa de seleção é desnecessária. Esta tarefa é descrita com mais detalhes na Seção 2.2.

Na etapa de combinação ou integração (Figura 1-3), uma decisão final é tomada tendo em vista as opiniões obtidas pelos especialistas, conforme uma regra de combinação. Essa etapa pode não ser necessária caso apenas um classificador tenha sido selecionado. A combinação é descrita com mais detalhes na Seção 2.3.

2.1 Geração de classificadores

A geração é o primeiro passo do processo de um SMC. Seu objetivo é criar um *pool* de classificadores preciso e diversificado. Para isso, podem ser gerados conjuntos de dados diferentes, a partir de um conjunto de dados original, ou usando variações dos parâmetros dos classificadores, ou ainda, utilizando diferentes tipos de algoritmos de aprendizagem.

O *pool* de classificadores pode ser gerado por métodos homogêneos ou heterogêneos. Os primeiros utilizam algoritmos do mesmo tipo no processo de geração, variando os parâmetros do método de aprendizagem ou o conjunto de dados. Já os heterogêneos utilizam algoritmos diferentes no treinamento dos classificadores. As técnicas homogêneas mais populares e mais aplicadas são o *Bagging* (BREIMAN, 1996), *Boosting* (FREUND; SCHAPIRE, 1996) e *Random Subspaces*(HO, 1998). Tais estratégias são detalhadas nas seções seguintes.

2.1.1 *Bagging*

O *Bagging* é um método proposto por Breiman (1996) que consiste em gerar, a partir de um conjunto de dados original, vários subconjuntos de dados distintos, de forma aleatória, usando re-amostragem com reposição e mantendo a proporção entre as classes. Espera-se que, com a variação dos elementos dos subconjuntos, a diversidade seja alcançada.

A Figura 2 apresenta um exemplo de funcionamento do *Bagging*. Nela, é apresentado um conjunto de dados de treino X_{train} composto de dez instâncias, das quais seis pertencem à classe a $(a_1, a_2, a_3, a_4, a_5, a_6)$ e quatro à classe b (b_1, b_2, b_3, b_4). Após aplicação do método com re-amostragem de 50% do tamanho do conjunto de treino, são gerados M subconjuntos $S_1, S_2, ..., S_M$. Podemos observar que os subconjuntos gerados mantiveram a proporção das classes do conjunto de dados original, em que cada um foi composto por três elementos da classe a e dois elementos da classe b. Cada subconjunto $S_1, S_2, ..., S_M$ gerado é empregado no treinamento de um classificador $C_1, C_2, ..., C_M$.

2.1.2 *Boosting*

No *Boosting*, de forma semelhante ao *Bagging*, cada classificador é treinado usando um conjunto de dados diferente. Porém, a principal diferença em relação ao *Bagging* é que os conjuntos de dados re-amostrados são construídos especificamente para gerar aprendizados complementares, nos quais cada elemento busca aprender os erros cometidos pelo classificador anterior, dentro da sequência de classificadores que vão sendo criados (FREUND; SCHAPIRE, 1996).

O processo consiste em construir diversos classificadores de forma iterativa, variando

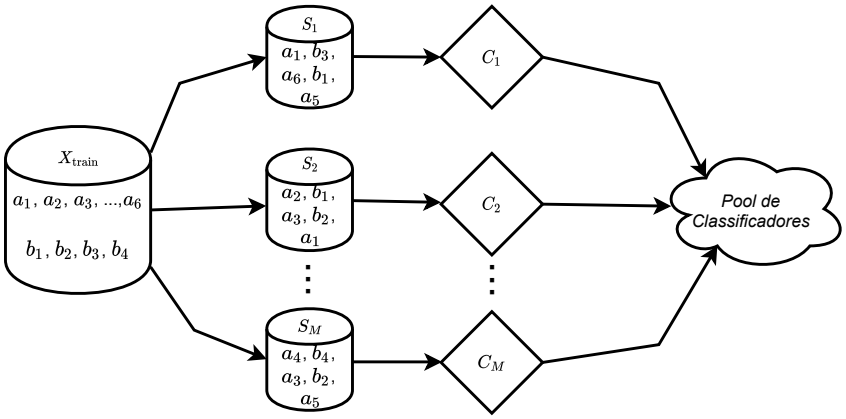


Figura 2 – Ilustração do método *Bagging* para geração de subconjuntos

o peso das instâncias do conjunto de treino. Inicialmente é gerado um conjunto de dados, no qual todas as amostras têm o mesmo peso. O conjunto formado é, então, submetido a um processo de classificação. Os elementos que forem classificados incorretamente terão seus pesos aumentados para que, na próxima iteração, tenham maior probabilidade de compor o conjunto de dados de forma que os classificadores seguintes se concentrem nos casos mais difíceis (FREUND; SCHAPIRE, 1996).

A Figura 3 ilustra o funcionamento da abordagem. O conjunto de treinamento X_{train} é a entrada para cada etapa na interação. Esta dependência ocorre devido à atualização dos pesos de cada amostra. Na ilustração, percebe-se que, a partir do conjunto de treino original (em que todas as instâncias têm o mesmo peso), é construído o subconjunto S_1. As instâncias que compõem este conjunto são, então, submetidas a um processo de classificação. Aquelas que forem classificadas erroneamente terão seus pesos aumentados para a iteração seguinte (formação do subconjunto S_2). Esse processo é realizado até que M subconjuntos sejam gerados. Assim como ocorre no *Bagging*, cada subconjunto $S_1, S_2, ..., S_M$ gerado é empregado no treinamento de um classificador $C_1, C_2, ..., C_M$.

O algoritmo de *Boosting* mantém a estratificação dos dados em cada um dos subconjuntos de forma a não mudar as proporções entre as classes presentes no conjunto de treino X_{train}. Além disso, na etapa de avaliação realizada em cada iteração do algoritmo, adota-se sempre o mesmo classificador.

2.1.3 *Random subspaces*

O *Random Subspaces* (RSS) é um método que gera os subconjuntos com base na seleção aleatória de um determinado número de atributos (características) sem repetição. Os atributos podem estar presentes em mais de um subconjunto, mas não podem se repetir em um mesmo subconjunto (HO, 1998).

Em cenários em que um o conjunto de dados do problema de classificação é formado por um número alto de atributos e com a possibilidade de que estes tenham características

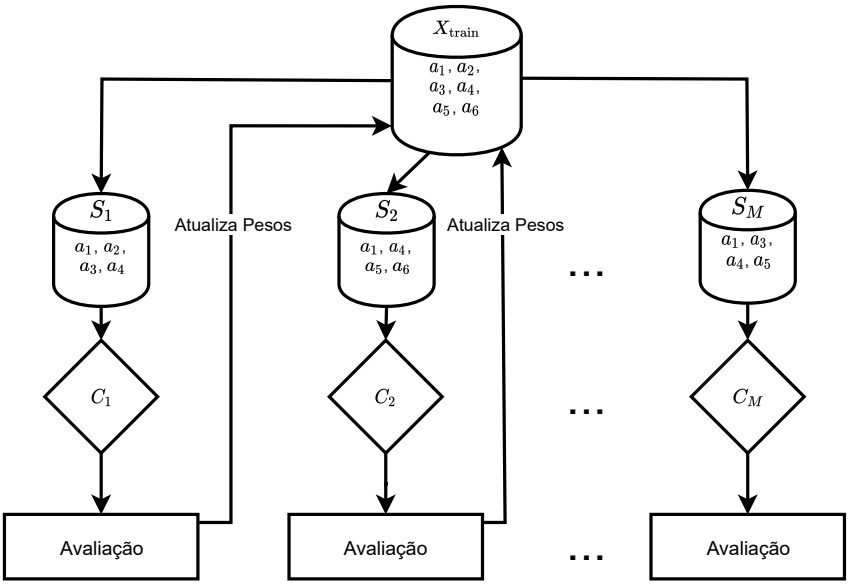


Figura 3 – Ilustração do método *Boosting* para geração de subconjuntos

parecidas, o método facilita a aprendizagem do classificador e consegue evitar a 'maldição da dimensionalidade' (HO, 1998; KUNCHEVA et al., 2001; PONTI JR., 2011).

O termo 'maldição da dimensionalidade' se refere a vários fenômenos que podem ocorrer na análise de dados em espaços com muitas dimensões (atributos), muitas vezes com centenas ou milhares de dimensões. Aprender a partir de um espaço de característica de alta dimensionalidade requer uma quantidade enorme de dados de treinamento para garantir que haja várias amostras com cada combinação de valores. Tal necessidade pode incorrer no gasto de um tempo considerável ou até mesmo levar ao esgotamento dos recursos computacionais. Outro ponto é que nem todos os atributos podem ser relevantes, muitos dos quais não implicando em melhoria no desempenho do classificador (BELLMAN, 2015).

A Figura 4 apresenta um exemplo do processo do *Random Subspaces*, a saber: com base no conjunto de treino X_{train} em que cada instância é formada por d atributos $(f_1, f_2, f_3, ..., f_d)$, dos quais são sorteadas quatro, sem repetição, para compor cada um dos M subconjuntos $S_1, S_2, ..., S_M$, que serão empregados no treinamento dos classificadores $C_1, C_2, ..., C_M$.

2.1.4 Geração direcionada por complexidade

Na pesquisa de Macià, Orriols-Puig e Bernadó-Mansilla (2010) relatou-se que os problemas reais disponíveis não cobrem todo o espaço da complexidade e, por isso, não possibilitam testar completamente o comportamento dos classificadores na região de fronteira.

Assim sendo, segundo os autores, surge a necessidade de desenvolver um cenário

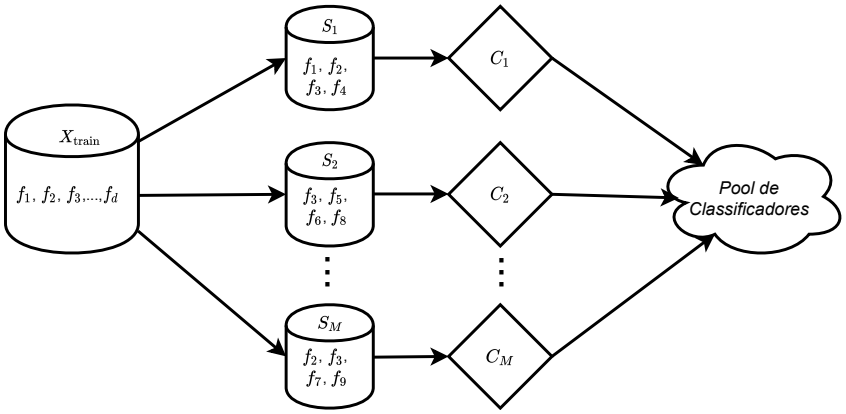


Figura 4 – Ilustração do método RSS para geração de subconjuntos

de treino mais apropriado. Eles propõem o uso de seleção de instâncias baseada em uma técnica evolutiva multiobjetiva (*Multi-Objective Genetic Algorithms*) para gerar conjuntos de dados que atendam a características específicas estabelecidas por descritores de complexidade. Nesse sentido, o algoritmo cria novas amostras sintéticas a partir das amostras reais do problema. Estas novas instâncias artificiais permitem oferecer uma cobertura mais completa do espaço de complexidade do problema.

No trabalho de Brun et al. (2018) um novo método de geração de *pools* de classificadores foi proposto com o objetivo de também formar conjuntos de instâncias em que seus descritores de complexidade estivessem mais bem dispersos no espaço, permitindo uma melhor representação das variações de complexidade dos problemas. O objetivo dos autores, além de aumentar a cobertura no espaço de complexidade, foi tentar formar um conjunto de classificadores que fosse acurado, de forma que quando adotados permitissem aumento de acurácia em comparação às estratégias tradicionais de geração de subconjuntos.

A solução proposta pelos autores empregou um algoritmo genético com função objetiva que combinava a acurácia dos classificadores treinados em cada subconjunto gerado e a dispersão dos conjuntos no espaço de complexidade. Os classificadores com maior acurácia sobre o conjunto de validação que estivessem mais distante dos demais no espaço de complexidade apresentariam melhor *fitting* - ou seja, de encaixe mais assertivo, apropriado. Dessa forma, a estratégia construída pôde obter *pools* com cobertura mais efetiva do espaço de complexidade e que, ao mesmo tempo, apresentaram melhor desempenho em termos de acurácia.

O trabalho de Monteiro et al. (2021) descreve um método de geração de *pool* de classificadores considerando a diversidade encontrada na complexidade dos dados e nas decisões do classificador. O método é composto por duas etapas.

Na primeira delas, é avaliado o comportamento das medidas de complexidade pertencente aos grupos de Medidas de Sobreposição e Medidas de Vizinhança, para cada subconjunto gerado a partir de um conjunto de treinamento. Então, é determinado quais

medidas devem ser usadas de acordo com sua variabilidade, sendo que as medidas com variabilidade mais alta são selecionadas.

Na segunda etapa, é realizada a geração do *pool* de classificadores com base em um algoritmo evolutivo multiobjetivo (*Multi-Objective Genetic Algorithms*), que otimiza a diversidade na complexidade e no espaço de decisão. Ele considera que cada subconjunto de dados é um cromossomo e cada gene é uma instância. O operador de *crossover* gera um novo indivíduo (subconjunto de dados) trocando instâncias de dois subconjuntos aleatórios. O processo de otimização explora a diversidade no espaço de complexidade usando as métricas selecionadas na primeira etapa do método. O objetivo é maximizar a distância média (dispersão) entre os subconjuntos de dados, considerando cada métrica de complexidade calculada em pares. Também visa maximizar a diversidade no espaço de decisão usando a métrica de Falta Dupla, calculada nas decisões dos classificadores treinados em cada subconjunto de dados, também realizada em pares.

2.1.5 Diversidade entre classificadores

A diversidade, que estima o grau de concordância e discordância entre classificadores, é considerada uma característica essencial entre os componentes de um *pool*. Um *ensemble* somente pode ter um desempenho melhor do que o melhor classificador individual, quando existir diversidade entre os classificadores do conjunto (SHIPP; KUNCHEVA, 2002; KUNCHEVA; WHITAKER, 2003; BROWN et al., 2005). No entanto, devido à complexidade de interpretação da diversidade, não há, até então, consenso com referência a seu grau de influência na acurácia dos métodos. Um entendimento comum é que, quando os classificadores cometem erros estatisticamente independentes, a sua combinação tem possibilidade de levar a um aumento no desempenho do sistema (PONTI JR., 2011).

Diversas métricas para medir a diversidade foram propostas na busca por um entendimento aprofundado do seu papel na tarefa de classificação e para calcular a relação entre a acurácia e a diversidade do conjunto (SHIPP; KUNCHEVA, 2002; KUNCHEVA; WHITAKER, 2003). As medidas Estatística Q, Correlação, Falta Dupla e Discordância são aquelas que estimam a diversidade entre pares de classificadores. Por outro lado, as medidas de Entropia, Índice de Dificuldade, Variância de Kohavi-Wolpert, Relação de Concordância entre Classificadores, Diversidade Generalizada e a Diversidade de Erros Coincidentes consideram a opinião de um conjunto de classificadores simultaneamente (KUNCHEVA; WHITAKER, 2003).

2.2 Seleção de classificadores

A seleção é a segunda etapa no processo de um SMC. Nela, seleciona-se, a partir de um *pool*, o classificador, ou conjunto de classificadores, mais competente(s) para prever

a classe de cada amostra de teste. Os classificadores considerados mais competentes são aqueles cujas performances são melhores segundo um critério pré-definido. Essa etapa é opcional, pois não é utilizada por alguns métodos de seleção, como por exemplo, quando é feita a combinação de todos os elementos presentes no *pool*.

A seleção de um ou mais classificadores pode ser realizada de forma estática ou dinâmica (GUNES et al., 2003; KO; SABOURIN; BRITTO JR., 2008). Na primeira, um subconjunto de classificadores é selecionado durante o treinamento, de acordo com um critério de competência estimado no conjunto de dados de validação. O mesmo conjunto de classificadores é usado para prever a classe de todas as amostras de teste.

Por outro lado, nas técnicas de seleção dinâmica, um ou mais classificadores são selecionados de acordo com as características de cada amostra de teste a ser classificada. A competência de cada classificador pode ser determinada por diversas formas. As mais comuns envolvem analisar a região de vizinhança da instância de teste no espaço de características ou informações de opinião dos elementos do *pool* sobre ela (GUNES et al., 2003; KO; SABOURIN; BRITTO JR., 2008; BRITTO JR.; SABOURIN; OLIVEIRA, 2014)

2.2.1 Seleção estática

A seleção estática de classificadores, representada na Figura 5, escolhe os melhores elementos de um *pool* C composto de M classificadores $(C_1, C_2, ..., C_M)$ para formar o subconjunto de classificadores C' de dimensão M' (em que $M' \leq M$) durante a fase de treinamento (ou offline), considerando como referência o conjunto de validação X_{val}. Pode-se perceber que o conjunto de classificadores C' não se altera durante a fase de teste, visto que os mesmos representantes escolhidos no treinamento são utilizados na classificação de todas as amostras de teste.

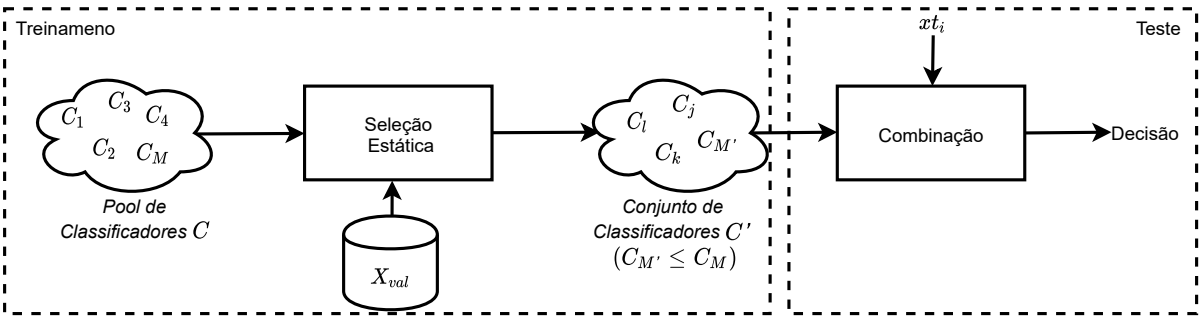


Figura 5 – Seleção estática de classificadores. Adaptado de (CRUZ; SABOURIN; CAVALCANTI, 2018)

2.2.2 Seleção dinâmica

A abordagem de seleção dinâmica pode escolher um único classificador (2.2.2.2), ou um ensamble (2.2.2.8). Exemplos de métodos desenvolvidos segundo as duas abordagens são apresentados nas seções seguintes.

A seleção dinâmica busca escolher o classificador (Figura 6), ou o conjunto de classificadores (Figura 7) mais competente, com base em um *pool* C composto de M classificadores $(C_1, C_2, ..., C_M)$, seguindo um determinado critério de seleção, para prever a classe de uma amostra de teste específica xt_i . Ela considera que cada classificador presente no *pool* é um especialista em uma região diferente do espaço de características (KO; SABOURIN; BRITTO JR., 2008). Na primeira estratégia (Figura 6), seleciona-se apenas o classificador C_j que ficará responsável por classificar a instância de teste xt_i. Já na estratégia de seleção de conjuntos (Figura 7), um grupo C' composto de M' (em que $M' \leq M$) classificadores $(C_j, C_k, C_l, C_{M'})$ é formado para atribuir a classe à instância de teste. Neste caso, faz-se necessária uma estratégia de combinação dos votos dos elementos do conjunto selecionado. Importante notar nas ilustrações que, para as duas estratégias, a seleção ocorre durante a fase de teste do processo.

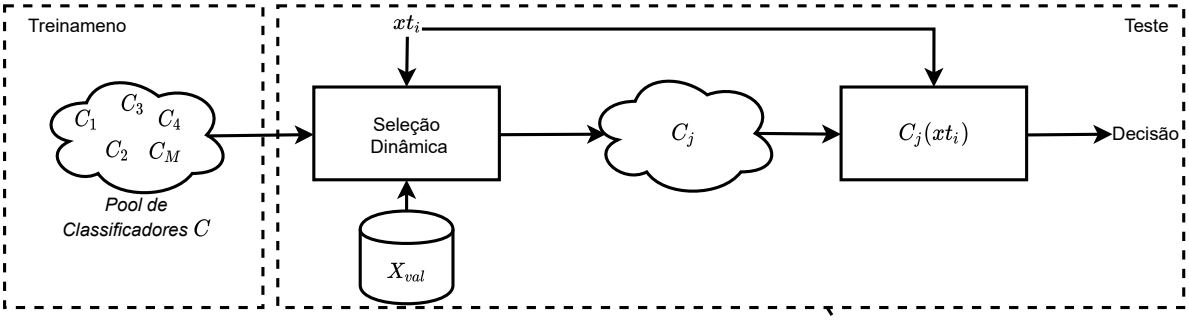


Figura 6 – Seleção dinâmica de um classificador. Adaptado de (CRUZ; SABOURIN; CAVALCANTI, 2018)

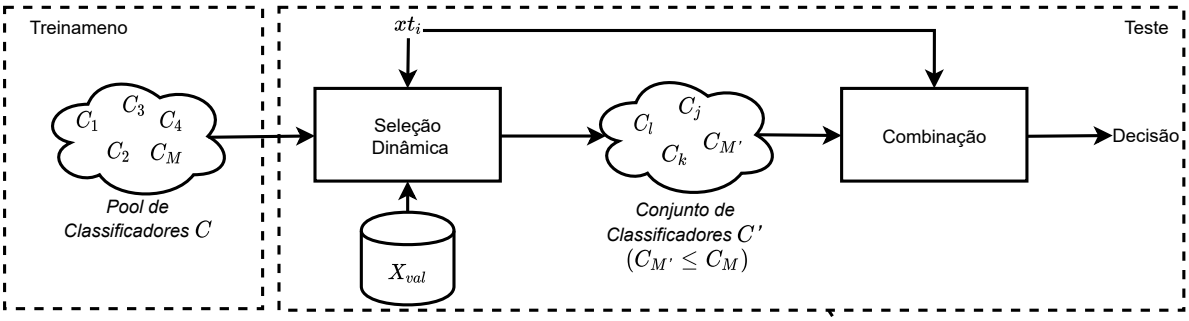


Figura 7 – Seleção dinâmica de um conjunto de classificadores. Adaptado de (CRUZ; SABOURIN; CAVALCANTI, 2018)

A competência do classificador define o nível de confiança que temos nos especialistas para uma determinada tarefa de classificação. A ideia de competência é amplamente

utilizada na área de conhecimento, a fim de escolher um dentre vários classificadores para melhor solução de um problema em questão (CRUZ et al., 2015).

O critério de seleção é o componente mais importante das técnicas de seleção dinâmica. Diferentes abordagens de seleção podem ser utilizadas para estimar a competência e avaliar o desempenho de cada classificador. Tais critérios são divididos em individual e de grupos. Na primeira estratégia, temos os métodos baseados em *ranking*, acurácia, probabilístico, relacionamento, meta-aprendizagem, oráculo e complexidade. Já nos critérios de seleção de grupos temos diversidade, ambiguidade e manipulação dos dados (BRITTO JR.; SABOURIN; OLIVEIRA, 2014) (CRUZ; SABOURIN; CAVALCANTI, 2018).

A competência de cada classificador normalmente é calculada considerando uma região local em torno da amostra de teste. Esta região é chamada de região de competência e é composta pelas instâncias vizinhas à amostra de teste xt_i no conjunto de validação X_{val}. Para demarcar essa região, podemos utilizar estratégias de vizinhanças, tais como o método dos vizinhos mais próximos (KNN) ou o algoritmo de agrupamento (k-means) (WOODS; KEGELMEYER; BOWYER, 1997) (KO; SABOURIN; BRITTO JR., 2008) (KUNCHEVA, 2000) (SOARES et al., 2006) (CRUZ; SABOURIN; CAVALCANTI, 2018).

Conjunto de validação (X_{val}) tem por objetivo calibrar os modelos, bem como extrair meta-informações, como os vizinhos de um determinado espaço de características dado. Para tanto, no momento da classificação o conjunto X_{val} deve estar disponível.

2.2.2.1 Oráculo

O Oráculo (do inglês *Oracle*) é um modelo abstrato definido por Kuncheva (2002) que, sempre que possível, seleciona do *pool* o classificador que previu a classe correta para uma dada amostra teste. Ou seja, se pelo menos um dos classificadores for capaz de opinar corretamente sobre a classe da instância de teste, ela é considerada um acerto.

Esse conceito é importante na literatura sobre os SMC's visto que serve como um teto para a precisão do conjunto de classificadores e é usado para determinar se os resultados obtidos pelas técnicas de seleção dinâmica propostas estão próximos da precisão ótima ou se ainda apresentam margem para melhora.

2.2.2.2 Seleção dinâmica do classificador

A seleção dinâmica de um único classificador busca identificar e selecionar de um *pool*, com base em algum critério pré-definido, o classificador mais competente para classificar cada nova amostra de teste. O classificador selecionado será responsável em atribuir classe para aquela amostra de teste específica. Para um novo padrão, o processo de seleção é realizado novamente.

Nas próximas seções, serão apresentadas alguns métodos de seleção dinâmica do classificador.

2.2.2.3 Seleção baseada em *ranking*

O método *Rank* foi um dos primeiros critérios de seleção propostos para estimar a competência dos classificadores na seleção dinâmica. A competência do classificador é estimada baseando-se no número de instâncias classificadas corretamente, de forma consecutiva, na região de competência do conjunto de validação. O classificador que classificar corretamente o maior número de instâncias seguidamente é selecionado (SABOURIN et al., 1993).

2.2.2.4 Acurácia local total - OLA

O método OLA (*Overall Local Accuracy*) utiliza a acurácia local como critério de seleção. Ele considera o número de instâncias classificadas corretamente na região de competência (K vizinhos mais próximos) do conjunto de validação. O classificador que tiver o maior percentual de acertos é selecionado para prever a classe da amostra de teste (WOODS; KEGELMEYER; BOWYER, 1997).

A Figura 8 apresenta um exemplo prático do funcionamento do método. Nesse cenário, temos cinco classificadores C_1, C_2, C_3, C_4 e C_5 que classificam sete vizinhos de uma amostra de teste xt_i (representados por $K_1, .., K_7$) em uma das classes possíveis $w = \{w_1, w_2\}$. Nas linhas, são representados os votos de cada classificador para cada vizinho de xt_i e, na última linha, a classe real de cada vizinho. Os elementos em negrito são os erros cometidos pelos classificadores. Na representação o classificador C_4 é selecionado por obter o maior percentual de acertos, estimando corretamente a classe de seis $(K_1, K_2, K_3, K_4, K_5, K_6)$ dos sete vizinhos de xt_i.

	K_1	K_2	K_3	K_4	K_5	K_6	K_7	
C_1	2	**1**	1	2	**2**	1	2	5/7 = 0,71
C_2	**1**	2	1	**1**	1	**2**	**1**	3/7 = 0,43
C_3	2	**1**	**2**	2	**2**	1	2	4/7 = 0,57
C_4	2	2	1	2	1	1	**1**	6/7 = 0,85
C_5	**1**	**1**	**2**	2	**2**	**2**	2	2/7 = 0,28
$Classe(w)$	2	2	1	2	1	1	2	

Figura 8 – Exemplo prático do método OLA (*Overall Local Accuracy*)

2.2.2.5 Acurácia local da classe - LCA

O método LCA (*Local Class Accuracy*) também usa a acurácia local como critério de seleção, mas com relação a alguma classe de saída. A competência do classificador é

estimada com base nas classes atribuídas pelos classificadores às instâncias na região de competência (Kvizinhos mais próximos) do conjunto de validação.

Considere um classificador que atribui, a uma amostra de teste xt_i, a classe w_i. Podemos determinar a porcentagem dos vizinhos na região (Kvizinhos mais próximos) que foram classificados corretamente com a classe w_i por este classificador. O elemento com percentual mais alto é selecionado (WOODS; KEGELMEYER; BOWYER, 1997) para atribuir a classe à instância de teste.

A Figura 9 ilustra o procedimento realizado pelo método. Nela, temos cinco classificadores C_1, C_2, C_3, C_4 e C_5, que classificam uma amostra de teste xt_i e seus sete vizinhos (representados por $K_1, .., K_7$) em uma das classes possíveis $w = \{w_1, w_2\}$. A opinião de cada classificador sobre a instância de teste é apresentada entre os parênteses na primeira coluna. O percentual de acerto é calculado considerando o número de vizinhos que foram classificados corretamente com a mesma classe, atribuído à amostra de teste xt_i por cada classificador. Os elementos em negrito são os erros cometidos. O classificador C_5 é selecionado por ter o maior percentual de acertos, uma vez que ele atribuiu a classe 2 à instância xt_i e atribuiu a mesma classe a cinco (K_1, K_2, K_4, K_5, K_7) dos sete vizinhos, acertando em quatro ocasiões (K_1, K_2, K_4, K_7).

	K_1	K_2	K_3	K_4	K_5	K_6	K_7	
$C_1(xt_i = 2)$	2	1	1	2	**2**	1	2	3/4 = 0,75
$C_2(xt_i = 1)$	**1**	2	1	**1**	1	1	**1**	3/6 = 0,50
$C_3(xt_i = 2)$	2	1	**2**	2	**2**	1	2	3/5 = 0,60
$C_4(xt_i = 1)$	**1**	**1**	2	**1**	1	1	**1**	2/6 = 0,33
$C_5(xt_i = 2)$	2	2	1	2	**2**	1	2	4/5 = 0.80
$Classe(w)$	2	2	1	2	1	1	2	

Figura 9 – Procedimento do método LCA (*Local Class Accuracy*)

2.2.2.6 Seleção a priori

Esse método, assim como o OLA, usa a probabilidade de classificação correta de cada classificador para os K vizinhos da amostra de teste xt_i. Essa estratégia estima a competência dos classificadores ao contar o número de vizinhos da região de competência que eles conseguem rotular corretamente. O método também pondera a influência de cada vizinho da amostra de teste de acordo com a distância euclidiana entre eles, o qual considera que os elementos mais próximos têm maior influência (GIACINTO; ROLI, 1999) (DIDACI et al., 2005).

2.2.2.7 Seleção a posteriori

Esse método usa a probabilidade de classificação correta de cada classificador para cada vizinho da amostra de teste considerando a classe de saída dessa amostra de teste.

Para isso, utiliza a acurácia do classificador.

Considere um classificador que atribui uma amostra de teste xt_i à classe w_i. Podemos determinar a porcentagem dos vizinhos na região de competência (K vizinhos mais próximos) que foram classificados corretamente com a classe w_i por este classificador (exatamente como o método LCA). Além disso, o método também pondera a influência de cada vizinho da amostra de teste de acordo com a distância euclidiana entre eles. Aqui, estabelece-se que os vizinhos mais próximos têm maior peso. (GIACINTO; ROLI, 1999) (DIDACI et al., 2005).

2.2.2.8 Seleção dinâmica de conjunto de classificadores

A seleção dinâmica de conjuntos de classificadores tem como finalidade escolher, de um *pool* de classificadores, um conjunto de classificadores especialistas para solução de um determinado problema.

Durante a classificação, podemos ter uma amostra de teste difícil de ser classificada por um determinado classificador que cobre uma certa região do espaço de características, visto que ela pode estar muito próxima da fronteira de decisão. Mas a mesma amostra pode ser facilmente reconhecida por outro classificador que cobre outra região do espaço de características, onde a amostra está longe da fronteira de decisão.

Nas próximas seções, serão apresentados alguns métodos de seleção dinâmica de conjunto de classificadores.

2.2.2.9 K oráculos mais próximos - KNORA

A técnica de seleção proposta por Ko, Sabourin e BRITTO JR. (2008) explora conceito do oráculo para conjuntos de classificadores. Ele é chamado de *KNORA (K-Nearest-Oracles)* e apresenta duas variações: o *KNORA-Eliminate (KN-E)* e o *KNORA-Union (KN-U)*.

No *KNORA-Eliminate (KN-E)*, são selecionados os classificadores que conseguem classificar corretamente todas as instâncias dos vizinhos de uma amostra na região de competência. Se nenhum classificador conseguir acertar todos os vizinhos, a exigência de vizinhos classificadores corretos é diminuída em uma unidade. Esse processo é realizado até que pelo menos um classificador consiga atender à exigência.

A Figura 10 apresenta o funcionamento do método. Nela, temos cinco classificadores C_1, C_2, C_3, C_4 e C_5 que classificam sete vizinhos $(K_1, .., K_7)$ de uma amostra de teste xt_i em uma das classes possíveis $w = \{1, 2\}$. Em cada linha, temos os votos dos classificadores sobre os vizinhos da instância de teste e, na última linha, é apresentada a classe correta de cada vizinho. Os elementos em negrito correspondem aos erros cometidos. No cenário representado, os classificadores C_1 e C_5 são selecionados por terem acertados todos os

vizinhos. Assim, as suas opiniões serão combinadas para se obter um voto único para a amostra de teste xt_i.

	K_1	K_2	K_3	K_4	K_5	K_6	K_7	
C_1	2	2	1	2	1	2	1	Sim
C_2	**1**	2	1	**1**	1	**1**	**2**	Não
C_3	2	1	**2**	2	**2**	2	1	Não
C_4	**1**	2	1	**1**	1	2	**1**	Não
C_5	2	2	1	2	1	2	1	Sim
$Classe(w)$	2	2	1	2	1	2	1	

Figura 10 – Ilustração do funcionamento do método *KNORA-Eliminate (KN-E)*

O *KNORA-Union (KN-U)* adota um critério de seleção menos severo. Nele, são selecionados os classificadores que conseguem classificar corretamente pelo menos um dos vizinhos de uma amostra de teste na região de competência.

A Figura 11 apresenta o funcionamento do método. A representação é similar àquela presente na Figura 10. Nela, temos cinco classificadores C_1, C_2, C_3, C_4 e C_5 que classificam sete vizinhos $(K_1, .., K_7)$ de uma amostra de teste xt_i para uma das possíveis classes $w = \{1, 2\}$. Os elementos em negrito são casos nos quais houve erro na classificação dos vizinhos. Os classificadores C_1, C_2, C_3 e C_4 são selecionados porque acertaram pelo menos um dos vizinhos de xt_i. Uma vez que a seleção envolveu mais de um classificador, alguma estratégia de combinação é necessária para se obter o voto final sobre o teste.

	K_1	K_2	K_3	K_4	K_5	K_6	K_7	
C_1	2	2	1	2	1	2	1	Sim
C_2	**1**	**1**	1	**1**	1	2	**2**	Sim
C_3	2	1	**2**	2	**2**	2	**2**	Sim
C_4	2	2	1	**1**	1	2	1	Sim
C_5	**1**	**1**	**2**	**1**	**2**	**1**	**2**	Não
$Classe(w)$	2	2	1	2	1	2	1	

Figura 11 – Ilustração do funcionamento do método *KNORA-Union (KN-U)*

2.2.2.10 Seleção baseada em diversidade e acurácia

A proposta desenvolvida por Santana et al. (2006) utiliza a acurácia e a diversidade dos classificadores do *pool* como critério de seleção para formar *ensembles* dinamicamente de acordo com a amostra de teste. Ela emprega dois procedimentos para definir a região de competência em um conjunto de validação. O primeiro usa um algoritmo de agrupamento (*K-means*) enquanto o segundo utiliza o método dos vizinhos mais próximos (KNN).

No procedimento de agrupamento, os dados do conjunto de validação são agrupados em um determinado número de grupos com o algoritmo *K-means*. Para cada *cluster* gerado,

os classificadores são ordenados em ordem decrescente de acurácia e ordem crescente de diversidade. Para ordenar os classificadores em uma ordem crescente de diversidade, qualquer medida de diversidade de pares pode ser usada. Nessa proposta, a medida de diversidade de Falta Dupla (KUNCHEVA; WHITAKER, 2003) é usada. No momento da classificação, cada amostra de teste é atribuída ao grupo que possuir o centroide mais próximo (distância euclidiana). Na sequência, selecionam-se os M classificadores com maior acurácia do grupo. Destes, são selecionado os M' classificadores com maior diversidade para compor os membros do *ensamble*. Com o conjunto de especialistas formado, a classe é atribuída à amostra de teste, segundo alguma regra de combinação.

O segundo procedimento diferencia-se do primeiro na definição da região de competência dos classificadores. Nesse procedimento, essa região é definida pelo método KNN. Para cada amostra de teste, são encontrados os K vizinhos mais próximos. Com base nessa vizinhança, os classificadores são ordenados em ordem decrescente de acurácia e ordem crescente de diversidade. As demais etapas seguem como o primeiro procedimento.

A medida de diversidade Falta Dupla de (KUNCHEVA; WHITAKER, 2003) é definida como a proporção de casos que foram classificados erroneamente por ambos os classificadores, conforme apresentado na Equação 2.1.

$$N_{i,k} = \frac{N^{00}}{N^{11} + N^{10} + N^{01} + N^{00}} \tag{2.1}$$

Onde: N^{00} é o número de instâncias que ambos classificadores classificados erroneamente; N^{01} é o número de instâncias que o primeiro classificador classificou corretamente e o segundo classificou incorretamente; N^{10} é o número de instâncias que o primeiro classificador classificou incorretamente e o segundo classificou corretamente e N^{11} é o número de instâncias que ambos os classificadores classificaram incorretamente.

2.2.2.11 Seleção baseada em diversidade - SDES

O método de Yan et al. (2013) utiliza a diversidade como critério na seleção dinâmica dos classificadores, apoiando-se na ideia de que quanto maior diversidade entre os classificadores selecionados, maior possibilidade de acerto na classificação das amostras. A seleção do conjunto de classificadores especialistas envolve a busca pelos vizinhos mais próximos para cada nova amostra de teste, o que pode ser demorado. Essa busca é evitada nessa proposta. O método inclui duas etapas e é chamado de *Sorting-Based Dynamic Classifier Ensemble Selection (SDES)*.

Na primeira etapa, os classificadores são ordenados pela sua diversidade em ordem decrescente. A diversidade entre dois classificadores é calculada usando a estatística kpav de Fleiss (KUNCHEVA; WHITAKER, 2003). Na segunda etapa, os classificadores são selecionados de acordo com a ordem estabelecida na primeira etapa até um limite de

confiança pré-estabelecido. Se a adição do novo classificador ao conjunto não leva o método a ultrapassar o limite de confiança estabelecido, o classificador não é escolhido. E, uma vez que mais de um classificador pode ser selecionado, nesse caso, uma estratégia de combinação por voto majoritário deve ser adotada.

2.3 Combinação de classificadores

A integração é a terceira e última etapa no processo de um SMC. A etapa de seleção pode ter como resultado um subconjunto composto por mais de um classificador, fazendo com que seja necessário combinar suas saídas a fim de tomar uma decisão única, a qual pode ser obtida através das regras de combinação. Segundo PONTI JR. (2011), ela pode ser realizada por três topologias diferentes, as quais se intitulam: paralela, serial e híbrida.

Na primeira, cada classificador escolhido apresenta sua opinião sobre a instância de forma independente. Essas respostas são direcionadas para um único lugar, que produzirá um resultado final, conforme uma certa regra de decisão (voto majoritário, por exemplo).

Na abordagem sequencial, os classificadores são arranjados de forma que, se o primeiro, ao tentar classificar a instância de teste, não alcançar um grau de confiança exigido, a instância é passada ao classificador seguinte. Esse processo é executado até que algum classificador atinja o patamar exigido ou até que todos tenham sido testados. Geralmente, os classificadores são ordenados do mais simples ao mais complexo, de maneira que, se o problema a ser classificado não for difícil, os primeiros classificadores conseguirão atribuir a classe ao teste, sem a necessidade de executar métodos mais complexos.

Na topologia híbrida, combinam-se as duas maneiras paralela e sequencial para otimizar o desempenho.

Em geral, as saídas produzidas por classificadores podem ser divididas em três categorias, as quais são ilustradas na (Figura 12).

- Abstrata: os classificadores produzem como saída apenas a classe escolhida;
- *Rank:* o classificador produz uma lista ordenada decrescente da classe mais provável para a menos provável;
- Probabilística: os classificadores geram uma lista das possíveis classes com as probabilidades ou *scores* relacionados a cada uma.

A combinação pode ser feita através de formas simples, como as regras da soma, do produto, do mínimo, do máximo, da média e da mediana. Podemos também combinar os classificadores através do voto da maioria, do voto ponderado ou usando métodos mais

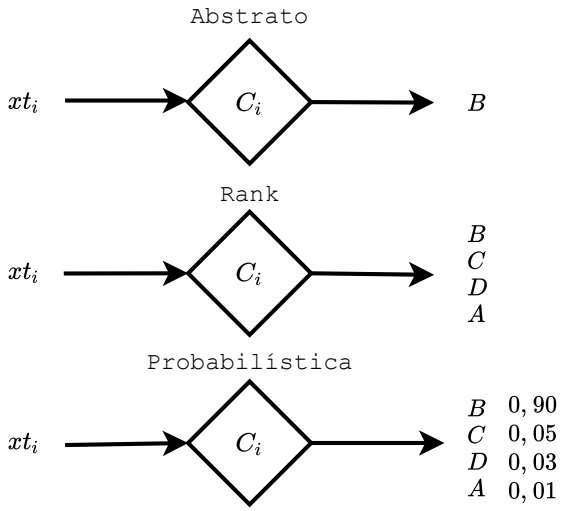


Figura 12 – Categorias de saídas dos classificadores em um SMC

avançados como o *Borda Count* (KITTLER, 1998; RUTA; GABRYS, 2000; GUNES et al., 2003; KUNCHEVA, 2014). As estratégias de combinação mais comuns na literatura são detalhadas nas seções seguintes.

2.3.1 Produto

A regra do produto é baseada na multiplicação dos níveis de confiança (probabilidades) fornecidos pelos classificadores para cada classe. A classe cujo produto resultante for o maior, é declarada vencedora. É uma regra bastante severa para combinar classificadores, uma vez que basta que um dos classificadores esteja em desacordo com a decisão dos outros, para que o resultado final seja penalizado. Essa abordagem costuma ser indicada para cenários críticos em que o sistema não pode cometer erros (KITTLER, 1998; GUNES et al., 2003).

2.3.2 Soma

A regra da soma é baseada no somatório dos níveis de confiança (probabilidades) fornecidos pelos classificadores para cada classe. Os níveis de confiança são somados para cada classe e aquela cuja soma resultante for a mais elevada, é declarada vencedora. Ela é menos severa que a regra do produto (KITTLER, 1998; GUNES et al., 2003).

2.3.3 Máximo

Para cada classe possível, seleciona-se o seu maior valor de confiança entre todos os classificadores. O processo é realizado para todas as classes e aquela que apresentar o maior valor será atribuída ao exemplar de teste (KITTLER, 1998; GUNES et al., 2003).

2.3.4 Mínimo

Para cada classe possível, seleciona-se o seu menor valor de confiança entre todos os classificadores. O processo é realizado para todas as classes e aquela que apresentar o maior valor será atribuída ao exemplar de teste (KITTLER, 1998; GUNES et al., 2003).

2.3.5 Média

Essa regra consiste em aplicar a média em todos os níveis de confiança (probabilidades) fornecidos pelos classificadores para cada classe. O resultado pode ser influenciado pela presença de *outliers*. A classe com a média mais alta é a vencedora (GUNES et al., 2003).

2.3.6 Mediana

A combinação por mediana consiste em aplicar a mediana aos níveis de confiança (probabilidades) fornecidos pelos classificadores para cada classe. Aquela que possuir o maior valor entre todos os classificadores será a selecionada (KITTLER, 1998; GUNES et al., 2003).

2.3.7 Cenário de aplicação das regras

A Figura 13 apresenta um cenário onde uma amostra xt_i é analisada por três classificadores C_1, C_2 e C_3 e pode ser classificada em cinco classes $w = \{1, 2, 3, 4, 5\}$ possíveis. Como resultado dessa análise, temos o nível de confiança do classificador para cada uma dessas classes $(w_1, w_2, w_3, w_4, w_5)$. A classe com maior grau de confiança é o voto daquele classificador.

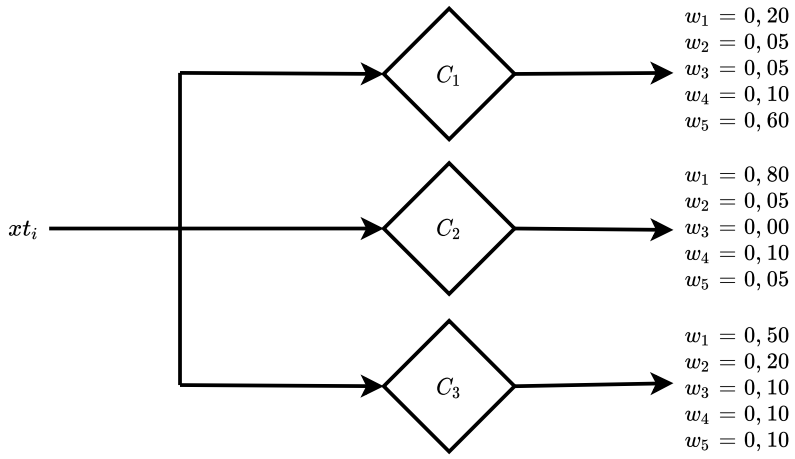


Figura 13 – Nível de confiança dos classificadores C_1, C_2 e C_3 de cada uma das w classes para amostra xt_i.

Tabela 1 – Combinação do nível de confiança dos classificadores

Regras de Combinação					
Classes / Regras	w_1	w_2	w_3	w_4	w_5
Produto	**0,0800**	0,0005	0,0000	0,0010	0,0030
Soma	**1,0000**	0,3000	0,1500	0,3000	0,7500
Max	**0,8000**	0,2000	0,1000	0,1000	0,6000
Min	**0,2000**	0,0500	0,0000	0,1000	0,0500
Média	**0,5000**	0,0100	0,0500	0,1000	0,2500
Mediana	**0,5000**	0,0500	0,0500	0,1000	0,1000
Voto Majoritário	**2**	0	0	0	1

A Tabela 1 ilustra as regras de combinação apresentadas acima. Ela apresenta a combinação do nível de confiança dos classificadores e determina a classe w_1 vencedora para a amostra xt_i.

2.3.8 Voto majoritário

Essa regra é similar ao processo eleitoral. A opinião de cada classificador sobre uma amostra de teste xt_i é contabilizada na forma de voto. Ela tem três padrões de consenso: unanimidade, maioria simples e pluralidade. No consenso de unanimidade, todos os classificadores devem dar a mesma resposta. Já na maioria simples, a classe vencedora deverá ter 50%+1 dos votos. E no consenso de pluralidade, a classe que tiver mais votos vence. As três abordagens são ilustradas na Figura 14. Suponha que as três formas correspondam às classes possíveis e que os tomadores de decisão sejam os dez classificadores individuais $C_1, ..., C_{10}$ do conjunto. A classe final em todos os três padrões de consenso seria a classe representada pelo quadrado (KITTLER, 1998; GUNES et al., 2003; KUNCHEVA, 2014).

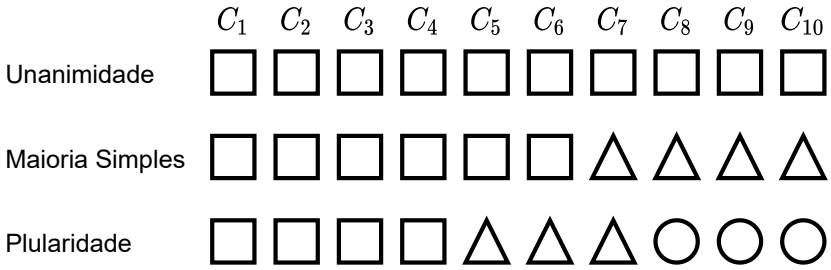


Figura 14 – Voto majoritário adaptado de (KUNCHEVA, 2014)

2.3.9 Voto majoritário ponderado

Essa regra, em vez de apenas somar número de votos, leva em conta as probabilidades de cada classe e um peso para cada classificador (KUNCHEVA, 2014).

As saídas dos classificadores devem fornecer uma lista ordenada com as possíveis classes e as suas devidas probabilidades. Os pesos dos classificadores são determinados de

acordo com seus desempenhos de classificação sobre o conjunto de validação. A ideia é dar maior importância (peso maior) para a opinião do classificador com maior exatidão.

As probabilidades de cada classe são multiplicadas pelo peso do classificador e seus valores somados. A classe final é então derivada da classe com a probabilidade mais alta. A Figura 15 apresenta o funcionamento da regra. Na ilustração, cada um dos classificadores $C_1, ..., C_5$ tem um peso atrelado W conforme seu desempenho na classificação do conjunto de validação. Os valores dos pesos são, então, multiplicados pela confiança de cada classificador em cada uma das três classes possíveis $w = \{1, 2, 3\}$. No exemplo retratado, a classe 2 é a que apresentou o maior somatório e, portanto, é a saída do sistema.

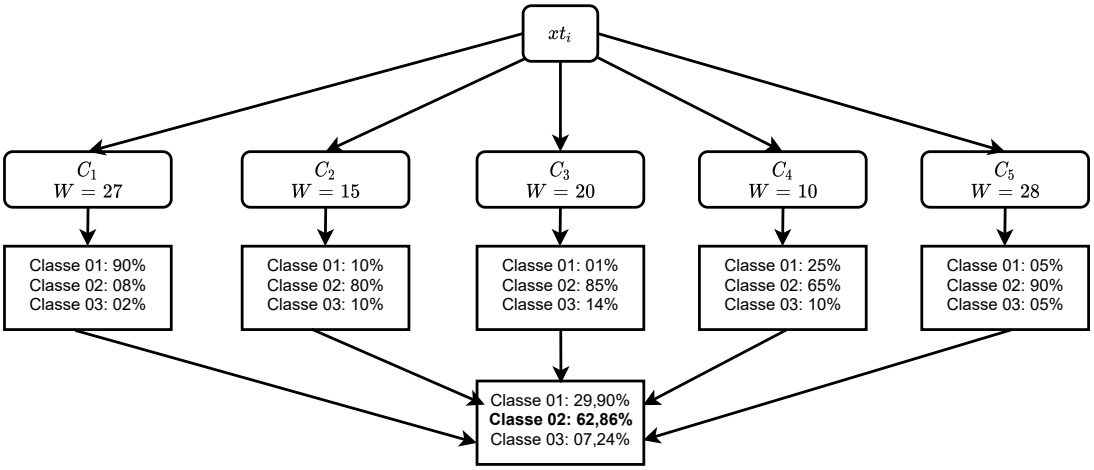


Figura 15 – Ilustração do funcionamento do voto majoritário ponderado

2.3.10 *Borda count*

Essa regra precisa que as saídas dos classificadores forneçam uma lista ordenada com as possíveis classes, sendo que também há a possibilidade de ser fornecida as probabilidades *ranking* para cada uma destas classes. Para cada posição da lista, é determinado um valor ou uma pontuação, que serão atribuídos à classe em conformidade com o índice de certeza de cada classe sugerida pelo classificador.

A combinação ocorre pela soma dos pontos e a classe que acumular mais pontos será a vencedora. A classe correta não necessariamente precisa aparecer como a primeira do *ranking,* mas apenas estar em posições próximas à primeira, na maioria das saídas dos classificadores (RUTA; GABRYS, 2000).

A Tabela 2 apresenta um exemplo quanto à aplicação desta regra, em que três classificadores C_1, C_2 e C_3 (sem probabilidades) opinam sobre uma amostra de teste xt_i, as possíveis saídas são cinco classes $w = \{1, 2, 3, 4, 5\}$. Uma pontuação entre um e cinco é atribuída para cada classe, de acordo com o grau de certeza do classificador. Para a classe com o menor grau de certeza, o classificador atribui o valor 1, à segunda classe com menor grau, o valor 2 é atribuído e, assim, sucessivamente. Esse processo é feito para todos os

classificadores do conjunto. Ao final do processo, somam-se os valores de todas as classes. Aquela que possuir o maior somatório, será a escolhida. No caso ilustrado na tabela, a classe 3 é a vencedora.

Tabela 2 – Funcionamento da regra *Borda Count*

C_1		C_2		C_3		Resultado	
Classe	Rank	Classe	Rank	Classe	Rank	Classe	Rank
3	**5**	4	5	1	5	**3**	**13**
4	4	**3**	**4**	**3**	**4**	4	11
2	3	5	3	2	3	1	9
1	2	1	2	4	2	2	7
5	1	2	1	5	1	5	5

2.3.11 *Ranking*

Nessa regra, atribui-se um ranque crescente para cada classe de acordo com a sua probabilidade, para cada classificador. A classe com maior certeza recebe ranque 1; a segunda classe, ranque 2; e, assim, sucessivamente. A classe escolhida será aquela que apresentar o melhor ranque. No caso de empate, por exemplo, de duas classes receberem o mesmo ranque, pode-se escolher a classe com o melhor ranque médio (RUTA; GABRYS, 2000).

2.4 Considerações finais

No presente capítulo, foram apresentados conceitos de classificação, bem como o funcionamento de sistemas de múltiplos classificadores. Esses sistemas atuam com vários classificadores, com objetivo de melhorar o desempenho no processo de reconhecimento perante sistemas individuais.

Nesse sentido, o funcionamento de SMC foi detalhado, desde a etapa da geração dos classificadores com suas principais estratégias de formação do *pool*, introduzindo a seleção dos classificadores, detalhando seus principais métodos e finalizando com a combinação dos classificadores.

A arquitetura escolhida, dentre as várias possíveis dentro de SMC, tem influência no desempenho do sistema. No entanto, outro aspecto importante que também influencia no processo de reconhecimento é o conjunto de dados que está sendo analisado. Diversas pesquisas (SÁNCHEZ; MOLLINEDA; SOTOCA, 2007; OKUN; VALENTINI, 2008; BRITTO JR.; SABOURIN; OLIVEIRA, 2014; BRUN et al., 2016; BRUN et al., 2018) apontam que o desempenho de um conjunto de classificadores e os métodos de seleção dinâmica sofrem influência da complexidade dos dados.

Os pesquisadores Ho e Basu (2002) e Macià, Orriols-Puig e Bernadó-Mansilla (2010) sugerem que, para se resolver um problema de classificação, compreender melhor a complexidade dos dados seja uma boa abordagem para a seleção de classificadores.

De forma a obter um entendimento maior sobre o comportamento dos dados, no capítulo seguinte serão apresentadas diversas medidas de complexidade, propostas na literatura, as quais explicam a relação dos dados e como o comportamento deles tem interferência direta no nosso objeto de estudo.

3 Medidas de Complexidade

As medidas de complexidade têm o objetivo de exprimir quão sobrepostas diferentes classes estão, o nível de complicação na região de fronteira entre as classes, ou mesmo, a distribuição das instâncias das classes no espaço de características. Em sua pesquisa, Ho e Basu (2002) propuseram a adoção de medidas de complexidade dos dados para tentar estimar o grau de dificuldade da tarefa de classificação.

Nas pesquisas de Ho e Basu (2002) e Sotoca, Sánchez e Mollineda (2005), as métricas de complexidade são divididas em três grupos principais: medidas de sobreposição de classes, medidas de separabilidade das classes e medidas de geometria, topologia e densidade. Entretanto, em pesquisas mais recentes, Lorena et al. (2018) e Lorena et al. (2019), realizaram uma abrangente revisão das medidas de complexidade e as dividiram em uma nova taxonomia composta de mais grupos como: medidas baseadas em características, medidas de linearidade, medidas de vizinhança, medidas de dimensionalidade, medidas de desequilíbrio de classe, medidas de correlação de Atributos, entre outros. Essas medidas serão discutidas, a seguir, nesse trabalho.

Para entendimento das equações, definimos X como o conjunto de dados, que é composto por n instâncias, cada uma das quais é descrita por d atributos preditivos. Logo, cada item de dados x_i, onde $i = 1, ..., n$, corresponde a uma linha de X, enquanto cada coluna f_j, onde $j = 1, ..., d$ corresponde a cada atributo de x_i. Cada item de dados também possui uma classe w_i e a composição de todas as classes formam um vetor w.

3.1 Medidas baseadas em atributos

Essas medidas concentram-se, principalmente, em verificar a sobreposição de diferentes classes no espaço de características e, para isso, avaliam o quão discriminativo são os atributos, analisando-se o intervalo e a distribuição de valores no conjunto de dados em relação a cada atributo e verificando sobreposições entre diferentes classes. Em muitas delas, cada atributo é avaliado individualmente. O problema pode ser considerado mais simples, caso haja pelo menos um atributo muito discriminativo. Por outro lado, o problema pode ser considerado mais complexo quando os atributos são mais sobrepostos.

3.1.1 Relação máxima do discriminante de fischer ($F1$)

Na sua versão original, essa medida calcula quão separadas são duas classes de acordo com um único atributo (HO; BASU, 2002). O cálculo é feito comparando-se as médias e **o desvio padrões** das classes para cada atributo específico. Ele é apresentado

na Equação 3.1, onde μ_1, μ_2, σ_1^2 e σ_2^2 representam as médias e os desvios padrões de duas classes representadas pelo número 1 e 2, para um determinado atributo i.

$$F1_i = \frac{(\mu_{1_i} - \mu_{2_i})^2}{{\sigma_{1_i}}^2 - {\sigma_{2_i}}^2} \tag{3.1}$$

Considerando-se que o conjunto de dados é composto por d atributos, o valor de F1 será o maior dentre todos eles, conforme descrito na Equação 3.2.

$$F1^* = argmax(F1_1, F1_2, F1_3, ..., F1_d) \tag{3.2}$$

Porém, a Equação 3.1 contempla problemas com apenas duas classes. Em um conjunto de dados que tem problemas com mais classes, utiliza-se a Equação 3.3 (MOLLINEDA; SáNCHEZ; SOTOCA, 2005).

$$F1 = \frac{\sum_{i=1}^{L} n_i * \delta(\mu, \mu_i)}{\sum_{i=1}^{L} \sum_{j=1}^{n_i} \delta(x_j^i, \mu_i)} \tag{3.3}$$

Onde: n_i indica o número de amostras da classe i; L corresponde ao número de classes presentes no problema; δ é alguma métrica de distância (normalmente a distância euclidiana); μ refere-se à média geral, enquanto μ_i corresponde à média da classe i; e x_j^i representa a amostra j pertencente à classe i.

No trabalho de Lorena et al. (2019), os autores levaram em consideração o inverso da formulação $F1$ original, conforme apresentado na Equação 3.4. Com isso, os valores F1 tornam-se limitados no intervalo [0, 1]. Nesse caso, valores mais altos indicam problemas mais complexos, em que nenhum atributo individual é capaz de discriminar as classes, enquanto valores baixos de $F1$ são obtidos se houver pelo menos um atributo de entrada altamente discriminativo no conjunto de dados.

$$F1 = \frac{1}{1 + \max(r_{f_1}, r_{f_2}, r_{f_3}, ..., r_{f_d})} \tag{3.4}$$

Onde: r_{f_i} é uma razão discriminante para cada atributo f_i, que é apresentado na Equação 3.5; e d corresponde ao número de atributos presentes.

$$r_{f_i} = \frac{\sum_{j=1}^{L} n_{w_j} (\mu_{w_j}^{f_i} - \mu^{f_i})^2}{\sum_{j=1}^{L} \sum_{l=1}^{n_{w_j}} (x_{li}^j - \mu_{w_j}^{f_i})^2} \tag{3.5}$$

Onde: n_{w_j} é o número de amostras da classe w_j; $\mu_{w_j}^{f_i}$ é a média do atributo f_i; μ^{f_i} é a média dos valores f_i para todas as classes; e x_{li}^j é o valor individual do atributo f_i para uma amostra da classe w_j.

3.1.2 Relação máxima do discriminante de fischer do vetor direcional ($F1v$)

Essa medida foi usada em Orriols-Puig, Macià e Ho (2010) como complemento à F1. Ele busca por um vetor orientado que possa separar amostras de duas classes diferentes e considera o critério de Fisher direcional definido em Malina (2001) e representado na Equação 3.6.

$$vF = \frac{v^t B v}{v^t P v} \tag{3.6}$$

Onde: v é o vetor direcional no qual os dados são projetados para maximizar a separação das classes; B é a matriz de dispersão entre classes; e P é a matriz de dispersão dentro das classes. v, B e P são definidos de acordo com as Equações 3.7, 3.8 e 3.9.

$$v = P^{-1}(\mu_{w_1} - \mu_{w_2}) \tag{3.7}$$

Onde: μ_{w_i} é o centroide (vetor médio) da classe w_i; e P^{-1} é o pseudo-inverso de P.

$$B = (\mu_{w_1} - \mu_{w_2})(\mu_{w_1} - \mu_{w_2})^t \tag{3.8}$$

$$P = p_{w_1} \sum_{w_1} + p_{w_2} \sum_{w_2} \tag{3.9}$$

Onde: p_{w_i} é a proporção de exemplos na classe w_i; e $\sum_{w_1}$ é a matriz de dispersão da classe w_i. Tomando a definição de vF, a medida $F1v$ é calculada pela Equação 3.10.

$$F1v = \frac{1}{1 + vF} \tag{3.10}$$

Os valores mais baixos em $F1v$ indicam problemas de classificação mais simples, um hiperplano linear que será capaz de separar a maioria. Essa medida pode ser bastante cara para calcular, devido à necessidade da pseudo-inversa da matriz de dispersão. E, assim como $F1$, é baseada na normalidade das distribuições das classes.

3.1.3 Volume da região de sobreposição ($F2$)

Essa medida calcula a sobreposição das distribuições dos valores dos atributos, definidas pelas amostras de cada classe. É possível determiná-la encontrando, para cada atributo f_j, seus valores mínimos e máximos, em cada uma das classes w. Em seguida, calcula-se o tamanho da distância normalizada da região que é a faixa de intervalo compartilhado (Min-Max) dividido pelo (Max-Min), que é todo o intervalo alcançado pelo

atributo, considerando os valores apresentados pelas classes, conforme representado na Figura 16 para o atributo 1 das classes A e B.

Para determinar a sobreposição total de duas classes w_1 e w_2, deve-se calcular $F2$ para todos os atributos e então multiplicá-las, como apresentada na Equação 3.11.

$$F2 = \prod_{j}^{d} \frac{sobreposicao(f_j)}{intervalo(f_j)} = \prod_{j}^{d} \frac{\max\{0, \min\max(f_j) - \max\min(f_j)\}}{\max\max(f_j) - \min\min(f_j)}, \tag{3.11}$$

Onde:

$$\begin{aligned}
\min\max(f_j) &= \min(\max(f_j^{w_1}), \max(f_j^{w_2})) \\
\max\min(f_j) &= \max(\min(f_j^{w_1}), \min(f_j^{w_2})) \\
\max\max(f_j) &= \max(\max(f_j^{w_1}), \max(f_j^{w_2})) \\
\min\min(f_j) &= \min(\min(f_j^{w_1}), \min(f_j^{w_2}))
\end{aligned}$$

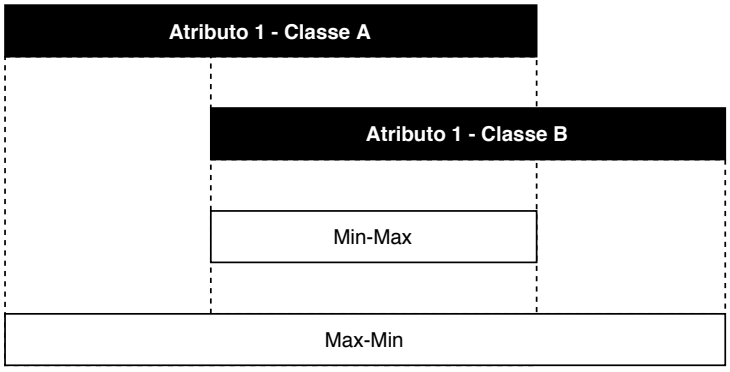


Figura 16 – Medida de Complexidade $F2$

Valores menores de $F2$ indicam uma menor sobreposição entre as classes e estão relacionados a problemas mais simples. Já valores mais altos de $F2$ indicam uma maior sobreposição entre as classes e estão relacionados a problemas mais complexos. Vale salientar que o valor de $F2$ deve ser zero quando houver pelo menos um atributo não sobreposto.

3.1.4 Eficiência máxima por atributo individual ($F3$)

Essa medida calcula a capacidade discriminativa individual de cada atributo na separação das classes e considera o valor máximo encontrado entre d atributos - no caso, o atributo que pode discriminar o maior número de amostra. Para cada atributo, é verificado se há sobreposição de valores entre amostras de classes diferentes e retorna-se a proporção do número de amostras que têm sobreposição pelo número total de amostras n do conjunto.

A razão discriminativa máxima é tomada como medida $F3$. Ela considera o número de amostras separadas pelo atributo mais discriminativo.

No trabalho de (LORENA et al., 2019), os autores fazem um complemento da medida $F3$, de tal maneira que valores mais altos sejam obtidos para problemas mais complexos. Para cada atributo, verifica-se se há sobreposição de valores entre amostras de classes diferentes. Se houver sobreposição, as classes são consideradas ambíguas nesta região. O problema pode ser considerado mais simples se houver pelo menos um atributo que apresente baixa ambiguidade entre as classes. A medida pode ser expressa pela Equação 3.12.

$$F3 = \min_{j=1}^{d} \frac{n_o(f_j)}{n} \tag{3.12}$$

Onde: $n_o(f_j)$ retorna o número de amostras que estão na região de sobreposição para o atributo f_j e pode ser expresso pela Equação 3.13. Valores baixos de $F3$, indicam problemas mais simples, nos quais poucas amostras se sobrepõem em pelo menos uma dimensão.

$$n_o(f_j) = \sum_{i=1}^{n} I(x_{ij} > \max\min(f_j) \wedge x_{ij} < \min\max(f_j)), \tag{3.13}$$

Onde: I é uma função que retorna 1 se seu argumento for verdadeiro e 0 caso contrário, enquanto maxmin (f_j) e minmax (f_j) são os mesmos definidos para $F2$ e x_{ij} corresponde ao valor do j-ésimo atributo da i-ésima amostra.

3.1.5 Eficiência coletiva dos atributos ($F4$)

Essa medida segue a mesma ideia apresentada na $F3$. No entanto, ela considera o poder discriminativo de todos os atributos em conjunto. Sendo assim, é considerada a eficiência coletiva dos atributos (ORRIOLS-PUIG; MACIà; HO, 2010).

O procedimento consiste em selecionar o atributo mais discriminativo, ou seja, aquele que apresentar a menor sobreposição entre classes diferentes. Em seguida, todas as amostras que podem ser separadas por esse atributo são removidas do conjunto de dados e o próximo atributo mais discriminativo é selecionado.

Esse procedimento é repetido até que todas as amostras sejam separadas ou que todos os atributos sejam analisados. Vale dizer que a medida retorna a proporção de instâncias que puderam ser discriminadas (ORRIOLS-PUIG; MACIà; HO, 2010).

A medida descrita aqui é diferente, em partes, da $F3$ que considera apenas o número de amostras separadas pelo atributo mais discriminativo, em vez de todos os atributos.

Assim, $F4$ fornece mais informações, visto que leva em consideração todos os atributos, pois destaca o poder discriminativo coletivo de todos eles.

No trabalho de Lorena et al. (2019), considerou-se como resultado de $F4$ a proporção de amostras que não foram discriminadas, conforme apresentado na Equação 3.14. A medida $F4$ é calculada depois de i repetições serem realizadas através do conjunto de dados, onde i está no intervalo $[1, d]$. Se o primeiro dos atributos de entrada já for capaz de discriminar todas as amostras do conjunto, i é 1, embora possa chegar a d no caso de todos os atributos terem que ser analisados.

$$F4 = \frac{n_o(f_{min}(X_i))}{n}, \tag{3.14}$$

Onde: $n_o(f_{min}(X_i))$ mede o número de instâncias na região de sobreposição do atributo f_{min} para o conjunto de dados da i-ésima repetição (X_i). Esse é o atributo mais discriminativo. Tomando a i-ésima iteração de $F4$, o atributo mais discriminativo no conjunto de dados (X_i) pode ser encontrado usando a Equação 3.15.

$$f_{min}(X_i) = \{f_j | \min_{j=1}^{d}(n_o(f_j))\}_{X_i}, \tag{3.15}$$

Onde: $n_o(f_j)$ pode ser encontrado pela Equação 3.13. E o conjunto de dados para cada interação pode ser definido pelas Equações 3.16 e 3.17.

$$X_1 = X \tag{3.16}$$

$$X_i = X_{i-1} - \{x_j | x_{ji} < \max\min(f_{min}(X_{i-1})) \vee x_{ji} > \min\max(f_{min}(X_{i-1}))\} \tag{3.17}$$

Ressalta-se que o conjunto de dados, na i-ésima iteração, é reduzido removendo-se todos os exemplares que puderam ser discriminados pelos atributos anteriores $f_{min}(X_{i-1})$.

3.2 Medidas de linearidade

As medidas de linearidade procuram avaliar até que ponto duas classes são separáveis linearmente, ou seja, se é possível separar as classes por um hiperplano.

3.2.1 Soma da distância de erro por programação linear ($L1$)

Essa medida avalia o quanto os dados são linearmente separáveis. Primeiro, constrói-se um classificador linear, de modo a diminuir os erros na separação das classes. Depois, a

medida pode ser calculada pela soma das distâncias de amostras classificadas incorretamente até o limite linear construído pelo classificador.

Um valor zero dessa medida indica que o problema é linearmente separável e pode ser considerado simples. Ao contrário, quanto maior o valor da medida, mais complexo é considerado o problema, no qual a separação linear não é eficiente.

As distâncias do erro das amostras erradas pode ser calculada somando os valores ε_i que corresponde à distância da instância até a fronteira definida pelo classificador para todas as instâncias classificadas erroneamente. A medida é ilustrada na Figura 17, em que as classes são representadas pelas diferentes figuras geométricas e a fronteira de classificação é representada pela reta entre os conjuntos.

O valor da medida do erro é obtido pela Equação 3.18. O valor $L1$ pode, então, ser calculado com a Equação 3.19.

$$SomaErroDist = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i \tag{3.18}$$

$$L1 = 1 - \frac{1}{1 + SomaErroDist} = \frac{SomaErroDist}{1 + SomaErroDist} \tag{3.19}$$

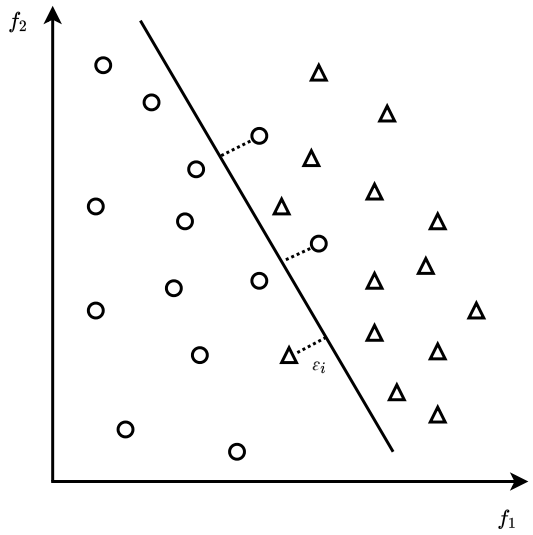


Figura 17 – Cenário em que um classificador linear erra ao classificar as três amostras

3.2.2 Taxa de erro do classificador linear ($L2$)

A medida $L2$ calcula a taxa de erro de um classificador linear. Ela avalia quantas amostras estão localizadas em regiões de classes diferentes da sua. O percentual de erro é obtido através da relação entre as amostras previstas incorretamente pelo total de amostras

do conjunto. Perceba que a Figura 17 ilustra um cenário para estimação da métrica. No caso, o valor de $L2$ seria $0,1111$ $(3/27)$. O cálculo é dado pela Equação 3.20.

$$L2 = \frac{\sum_{i=1}^{n} I(h(x_i) \neq w_i)}{n} \tag{3.20}$$

Onde: I é uma função que retorna 1 se seu argumento for verdadeiro e 0 caso contrário, $h(x_i)$ denota o classificador linear aplicado à i-ésima instância x_i; n representa o número de instâncias. A classe real de x_i é denotada por w_i.

3.2.3 Não linearidade de um classificador linear ($L3$)

A partir de um conjunto de dados de treino, cria-se um conjunto de teste por interpolação linear com coeficientes aleatórios entre pares de amostras escolhidas aleatoriamente da mesma classe.

Em seguida, um classificador linear é treinado com o conjunto de dados de treino e tem sua taxa de erro calculada (que corresponde ao valor de $L3$), aplicando-se o classificador sobre conjunto de dados de teste gerados, a partir da interpolação.

A $L3$ é aplicada em problemas de duas classes. Em cenários multi-classes, é necessário primeiro decompor o problema em subproblemas binários. A medida $L3$ pode ser expressa pela Equação 3.21. Os valores mais altos indicam uma maior complexidade.

$$L3 = \frac{1}{l} \sum_{i=1}^{l} I(h_X(x'_i) \neq w'_i) \tag{3.21}$$

Em que: I é uma função que retorna 1 se seu argumento for verdadeiro e 0 caso contrário; $h_X(x'_i)$ denota o classificador linear induzido a partir dos dados de treinamento originais X aplicado à i-ésima instância interpolada x'_i; l é o número instâncias interpoladas. A classe real de x'_i é denotada por w'_i.

3.3 Medidas de vizinhança

Essas medidas tentam encontrar a região de fronteira de decisão entre as classes e apontar a sobreposição, analisando as vizinhanças locais das instâncias.

3.3.1 A fração de pontos na região de fronteira ($N1$)

Nessa medida, primeiro é construída uma árvore geradora mínima *(MST - Minimum Spanning Tree)*, a partir do conjunto de dados de entrada, conforme ilustrado na Figura 18, em que cada vértice corresponde a uma amostra e as arestas são ponderadas de acordo

com a distância entre eles. Na ilustração, cada forma geométrica corresponde a uma classe distinta.

O valor de $N1$ é obtido calculando-se o percentual de vértices conectados a elementos de classes diferentes pelo total de vértices, conforme apresenta a Equação 3.22. Essas amostras estão próximas à fronteira das classes ou em áreas de sobreposição entre as classes. Valores mais altos de $N1$ indicam que há uma maior quantidade de sobreposição entre as classes na região da fronteira e, com isso, o problema de separação das classes é mais complexo.

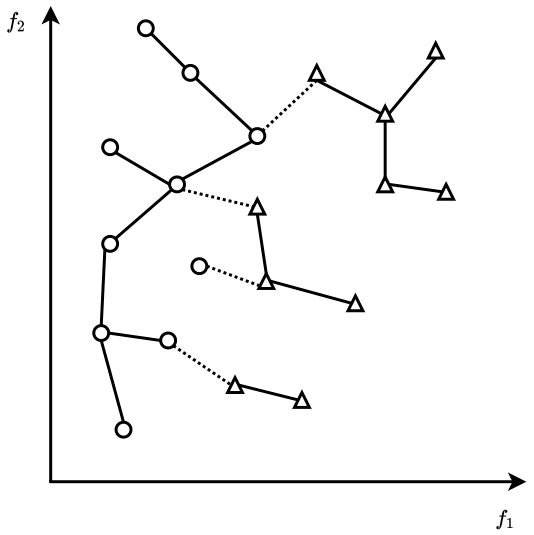


Figura 18 – Árvore geradora mínima construída com base no conjunto de dados de entrada

$$N1 = \frac{1}{n}\sum_{i=1}^{n} I((x_i, x_j) \in MST \wedge w_i, \neq w_j), \tag{3.22}$$

Onde: n é o número de vértices que compõem MST; I é uma função que retorna 1 se seu argumento for verdadeiro e 0 caso contrário; w_i e w_j são as classes de saída dos elementos x_i e x_j.

3.3.2 Proporção da distância intra/extra da classe do vizinho mais próximo ($N2$)

Essa medida divide as somas das distâncias entre cada amostra do conjunto de dados e seu vizinho mais próximo da mesma classe (intra-classe) pela soma das distâncias entre cada amostra do conjunto de dados e seu vizinho mais próximo de outra classe (extra-classe). O calculo da distância é realizado usando a distância euclidiana. A Figura 19 ilustra as distâncias intra e extra-classes para uma amostra específica de um conjunto de dados em que as figuras representam duas classes distintas. O cálculo é feito conforme Equação 3.23.

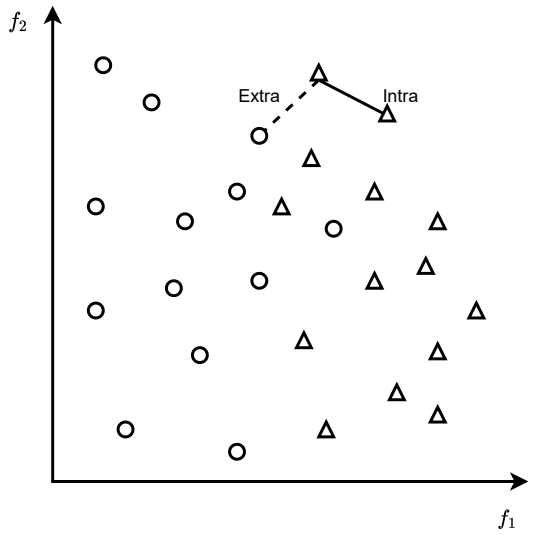


Figura 19 – Medida de complexidade $N2$

$$intra_extra = \frac{\sum_{i=1}^{n} \delta(x_i, NN(x_i) \in w_i)}{\sum_{i=1}^{n} \delta(x_i, NN(x_i) \in w_j \neq w_i)}, \tag{3.23}$$

Onde: $\delta(x_i, NN(x_i) \in w_i)$ corresponde à distância da instância x_i ao seu vizinho mais próximo (NN) dentro da sua própria classe w_i; e $\delta(x_i, NN(x_i) \in w_j \neq w_i)$ representa a distância de x_i para o vizinho mais próximo de outra classe $w_j \neq w_i$. O valor de N2 pode ser obtido conforme Equação 3.24.

$$N2 = 1 - \frac{1}{1 + intra_extra} = \frac{intra_extra}{1 + intra_extra} \tag{3.24}$$

3.3.3 Taxa de erro do classificador do vizinho mais próximo ($N3$)

A medida $N3$ corresponde à taxa de erro de um classificador $1NN$ *(one Nearest Neighbor)*. O percentual de erros é estimado pelo procedimento *leave-one-out*, que indica quão próximas são amostras de classes diferentes. Valores altos para $N3$ indicam que há sobreposição nas regiões de fronteiras, enquanto valores baixos indicam que existe um espaço entre os elementos da região da fronteira. A Equação 3.25 demonstra o cálculo da medida.

$$N3 = \frac{\sum_{i=1}^{n} (1NN(x_i) \neq w_i)}{n} \tag{3.25}$$

Onde: $NN(x_i)$ representa a previsão do classificador de vizinho mais próximo; n representa o número de instâncias; x_i é a i-ésima instância cuja classe é denotada por w_i.

3.3.4 Não linearidade do classificador do vizinho mais próximo ($N4$)

A ideia dessa medida é semelhante a $L3$. A diferença está no tipo do classificador adotado. Enquanto a medida $N4$ usa um classificador baseado em vizinhança (NN), a medida $L3$ emprega um classificador linear. O valor de $N4$ pode ser obtido pela Equação 3.26.

$$N4 = \frac{1}{l}\sum_{i=1}^{l} I(NN_X(x'_i) \neq w'_i) \tag{3.26}$$

Onde: $NN_X(x)$ denota o classificador de vizinhos mais próximos induzido a partir dos dados de treinamento originais X; l corresponde o número de instâncias interpoladas presentes no conjunto x'_i cujos elementos têm suas classes denotadas por w'_i.

Valores mais altos de $N4$ indicam problemas de maior complexidade. Outro ponto a observar em relação à medida $L3$, é que $N4$ pode ser aplicado diretamente a problemas de classificação com mais de duas classes, sem a necessidade de decompô-los em subproblemas binários.

3.3.5 Fração de esferas de cobertura máxima do dados ($T1$)

O processo para o cálculo de $T1$ consiste em primeiro construir circunferências com o centro posicionado em cada uma das amostras do conjunto de dados. O raio de cada círculo construído é aumentado progressivamente até que sua borda atinja uma amostra de outra classe. Depois do processo ser realizado para todas as amostras, os círculos menores que estiverem inteiramente contidos nos círculos maiores são eliminados. O valor de $T1$ corresponderá à razão entre o número dos círculos restantes, frente ao número total de amostras do conjunto (HO; BASU, 2002).

Uma implementação alternativa para $T1$ é proposta por Lorena et al. (2019). Essa implementação modifica a definição da interrupção no crescimento dos círculos. O raio de cada um pode ser determinado diretamente com base na matriz de distância entre todas as amostras. Após encontrar os raios de todas as circunferências, uma etapa posterior pode ser aplicada para verificar quais círculos serão eliminados - no caso, aquelas que estão dentro de círculos maiores.

3.4 Medidas de dimensionalidade

Esse grupo de medidas busca apresentar um indicativo da dispersão dos dados com base nas dimensões ocupadas pelos conjuntos. Quando temos bases de dados dispersas, a tendência é de enfrentar uma dificuldade maior de extrair bons classificadores. Essa

dificuldade ocorre, provavelmente, pela existência de regiões de baixa densidade que poderão ser classificadas aleatoriamente.

3.4.1 Número médio de pontos por dimensão ($T2$)

A medida $T2$ caracteriza a dispersão dos dados. Na sua versão original, divide-se o número de amostras n do conjunto de dados por sua dimensionalidade d (número de atributos do conjunto de dados) (HO; BASU, 2002).

No trabalho de Lorena et al. (2019), foi considerado o inverso da medida com finalidade de obter valores mais altos para conjuntos de dados mais complexos, conforme apresentado na Equação 3.27.

$$T2 = \frac{d}{n} \tag{3.27}$$

3.4.2 Número médio de dimensões PCA por pontos ($T3$)

Proposta por Lorena et al. (2012), a medida $T3$ é definida a partir de uma *Principal Component Analysis (PCA)* do conjunto de dados. O PCA busca uma transformação das características em funções lineares não correlacionadas que são capazes de descrever grande parte da variabilidade dos dados. Ele busca reduzir o número de dimensões de um conjunto de dados, projetando os dados em um novo plano.

Em vez da dimensionalidade bruta d do vetor de características original (como em $T2$), para se estimar $T3$, é usado o número de componentes principais necessários para representar 95% da variabilidade dos dados (d').Como a base de avaliação da dispersão dos dados. Obtido o valor de d', $T3$ é obtida através da Equação 3.28 em que n corresponde ao número de amostras do conjunto.

$$T3 = \frac{d'}{n} \tag{3.28}$$

3.4.3 Relação entre a dimensão PCA e a dimensão original ($T4$)

Essa medida fornece, segundo Lorena et al. (2012), uma estimação aproximada da proporção de dimensões relevantes para o conjunto de dados. Essa relevância é medida de acordo com o critério PCA. Como em $T3$, d' representa 95% da variabilidade dos dados e d corresponde à dimensão original. A medida pode ser calculada pela Equação 3.29.

$$T4 = \frac{d'}{d} \tag{3.29}$$

Quanto maior o valor de $T4$, mais atributos originais são necessários para descrever a variabilidade dos dados. Isso indica um relacionamento mais complexo das variáveis de entrada.

3.5 Medidas de balanço entre classes

Essas medidas tentam capturar perspectivas que podem influenciar amplamente o desempenho preditivo das técnicas de aprendizagem de máquina ao solucionar problemas de classificação de dados. O desbalanceamento de classes significa que há uma diferença significativa no número de instâncias por classe no conjunto de dados de treinamento. Quando essas diferenças existem no conjunto de dados, as técnicas de classificação de aprendizagem de máquina, em sua maioria, inclinam-se a favorecer a classe com maior número de instância e apresentam problemas de generalização. Caso o problema tenha um grande desequilíbrio na proporção de instância por classe, ele pode ser considerado mais complexo do que um problema para o qual as proporções são semelhantes.

3.5.1 Entropia da proporções das classe ($C1$ - $B1$)

A medida $B1$ determina o balanceamento das classes no conjunto de dados estimando-se a entropia normalizada da distribuição dos tamanhos de cada grupo, de acordo com a Equação 3.30.

$$B1 = -\frac{1}{log\left(n_{w_i}\right)} \sum_{i=1}^{L} p_{w_i} log(p_{w_i}) \tag{3.30}$$

Onde: L indica o número de classes do problema; n_{w_i} corresponde ao número de instâncias da classe w_i; e p_{w_i} consiste na proporção de instâncias da classe w_i dentro do conjunto.

Essa medida vai atingir o valor máximo para problemas balanceados, ou seja, problemas em que todas as classes têm proporções iguais. Estes podem ser considerados problemas mais simples, de acordo com o aspecto do balanceamento da classe.

3.5.2 Relação de balanço multi-classes ($C2$ - $B2$)

Essa medida calcula razão de desequilíbrio IR (Equação 3.32) para atender à proporção de todas as distribuições de classe (TANWANI; FAROOQ, 2010).

$$B2 = 1 - \frac{1}{IR} \tag{3.31}$$

Onde:

$$IR = \frac{L-1}{L} \sum_{i=1}^{L} \frac{n_{w_i}}{n - n_{w_i}} \tag{3.32}$$

Em que: n_{w_i} é o número de instâncias da i-ésima classe, L é o número de classes e n corresponde ao número total de instâncias. Os valores maiores de $B2$ são obtidos para problemas de desequilíbrio. Já o valor mínimo de $B2$ é alcançado para problemas totalmente balanceados.

3.6 Medidas de correlação de atributos

As medidas desse grupo consideram a correlação dos valores dos atributos com as saídas dos classificadores. Isso significa que se, pelo menos, um atributo estiver fortemente correlacionado à saída, tem-se um indicativo de problemas mais simples. As medidas desse grupo, em sua maioria, analisam cada atributo individualmente - o que se justifica pelo fato dos autores Lorena et al. (2018) estarem preocupados com medidas simples que forneçam um indicativo da complexidade do problema. Normalmente, uma combinação de mais medidas será necessária para caraterizar adequadamente o problema.

3.6.1 Correlação de atributos perante a saída ($C1$)

Nessa medida, segundo Lorena et al. (2018), primeiro, deve-se calcular o valor absoluto da correlação de Spearman entre cada atributo e todas as saídas. O valor absoluto é tomado porque ambos os extremos da medida de correlação, que variam entre [-1, 1], representam uma forte correlação, seja inversa ou direta. C1 representa o valor de correlação máxima obtido entre todos os atributos (semelhante à Razão Discriminante $F1$ de Fisher) perante as saídas, conforme representado na Equação 3.33.

$$C1 = \max_{j=1,\ldots,d} |\rho(f_j, w)| \tag{3.33}$$

Onde: ρ é a correlação de Spearman (SPEARMAN, 1904). Esta foi escolhida porque é não paramétrica e não assume nenhuma distribuição particular entre as saídas w (vetor de classes) e os vetores de atributos preditivos do conjunto de dados, f_j corresponde a cada atributo de um vetor de atributos x de d elementos. Valores mais altos de $C1$ indicam problemas mais simples, onde há pelo menos um atributo fortemente correlacionado com a saída. Por exemplo, para um conjunto de dados X_1, o valor $C1$ é 1, uma vez que o atributo f_1 tem correlação máxima com o atributo de saída(classe). Já para um conjunto de dados X_2, o valor $C1$ é 0,03, indicando que não há relacionamento de nenhum dos atributos preditivos com o atributo de saída (classe).

3.6.2 Correlação média de atributos perante a saída ($C2$)

A medida $C2$ tem funcionamento similar a $C1$: a diferença ocorre no fato de que, ao invés de selecionarmos apenas o valor máximo da correlação entre os atributos e saídas w (vetor de classes), ela calcula a média das correlações de todos os atributos com as saídas w (vetor de classes). Portanto, é uma medida aproximada da relação de todos os atributos com a saída (LORENA et al., 2018). Ela é definida na Equação 3.34.

$$C2 = \sum_{j=1}^{d} \frac{|\rho(f_j, w)|}{d}, \tag{3.34}$$

Onde ρ é a correlação de Spearman (SPEARMAN, 1904); w se refere às possíveis saídas (vetor de classes) e f_j corresponde a cada atributo de um vetor composto por d de atributos de cada elemento. Vale pontuar que valores mais altos da medida indicam problemas mais simples.

3.6.3 Eficiência de atributo individual ($C3$)

Essa medida calcula, para cada atributo, o número de instâncias que devem ser removidas do conjunto de dados até que um valor alto de correlação com a saída seja atingido. Em seu trabalho, Lorena et al. (2018) adotam $|\rho(f_j, w)| > 0,9$ como limiar para definir correlação alta, onde: ρ é a correlação de Spearman (SPEARMAN, 1904); w se refere a saídas (vetor de classes) e f_j corresponde a cada atributo de um vetor de atributos x de d elementos. Posteriormente, calcula-se o número de instância removidas pelo número total de instâncias.

A ideia é capturar o poder explicativo de cada um dos atributos, analisando quantas das instâncias originais podem se relacionar com a saída com uma confiança alta. Este cálculo é realizado para cada atributo individual e o mínimo dos valores encontrados é retornado. O resultado corresponde ao atributo mais relacionado à saída. Portanto, se um atributo preditivo de um conjunto de dados estiver altamente relacionado à saída, esta medida o interpreta como um problema de regressão simples. A medida $C3$ pode ser expressa na Equação 3.35.

$$C3 = \min_{j=1}^{d} \frac{n^j}{n} \tag{3.35}$$

Onde: n^j é o número de instâncias removidos para obter um alto valor de correlação entre o $j-$*ésimo* vetor de atributos e a saída; n refere-se ao número de instâncias do conjunto e d ao número de atributos.

3.6.4 Eficiência de atributos coletivos ($C4$)

Para se calcular essa medida, identifica-se, inicialmente, o atributo com maior correlação com a saída. Todas as instâncias com um pequeno valor residual ($|\varepsilon_i| \leq 0,1$), após um ajuste linear entre estes atributos e o atributo de saída, são removidas. Em seguida, o atributo mais correlacionado aos pontos de dados restantes é encontrado e o processo anterior é repetido até que todos os atributos tenham sido analisados ou nenhuma instância permaneça (LORENA et al., 2018). Diferente de C3, as instâncias removidas em uma rodada são desconsideradas nas etapas seguintes. A medida retorna a proporção de instâncias para os quais os valores absolutos dos resíduos do ajuste linear são maiores do que o limite 0,1 e é definida pela Equação 3.36.

$$C4 = \frac{\#\{x_i||\epsilon_i| > 0,1\}_{X_l}}{n'} \tag{3.36}$$

Onde: o numerador (n') fornece o número de instâncias que permanecem no conjunto de dados ao final da operação C4. X_l é o conjunto de dados a partir do qual esse número é calculado e l é o número de iterações realizadas pelo algoritmo ($l \in \{1, 2, ..., d\}$ em que d é o número de atributos) e x_i corresponde à i-ésima instância.

3.7 Demais medidas

Esta seção apresenta outras medidas encontradas na literatura que podem ser usadas para caracterizar a complexidade dos problemas de classificação.

3.7.1 Densidade ($D1$)

Essa medida pode ser definida pelo número médio de instâncias por unidade de volume onde todos os elementos estão distribuídos. O valor do volume é o produto dos comprimentos de todos os intervalos dos atributos, em que os valores são abrangidos por todas as classes (SOTOCA; SáNCHEZ; MOLLINEDA, 2005).

3.7.2 Volume da vizinhança local ($D2$)

Essa medida representa o volume médio ocupado pelos K vizinhos mais próximos de cada instância de treinamento (SÁNCHEZ; MOLLINEDA; SOTOCA, 2007). Seja $N_k(x_i)$ o conjunto dos K vizinhos mais próximos de uma amostra x_i de uma classe w_i, então o volume deste pode ser calculado pela Equação 3.37.

$$V_i = \prod_{h=1}^{d} (\max(f_h, N_k(x_i)) - \min(f_h, N_k(x_i))) \tag{3.37}$$

Onde: $\max(f_h, N_k(x_i))$ e $\min(f_h, N_k(x_i))$ representam, respectivamente, os valores máximos e mínimos do atributo f_h entre os K vizinhos mais próximos da amostra x_i.

Com valores de V calculados, o valor de D2 pode ser expresso como o valor médio de V_i para as n instâncias de treinamento, representado pela Equação 3.38.

$$D2 = \frac{1}{n}\sum_{i=1}^{n} V_i \tag{3.38}$$

3.7.3 Densidade da classe na região de sobreposição ($D3$)

As regiões de sobreposição possuem os casos mais críticos para a tarefa de classificação que originam a maior parte dos erros dos classificadores. Em vista disso, Sánchez, Mollineda e Sotoca (2007) propuseram a medida de densidade D3, que tem como objetivo determinar a densidade de cada classe nas regiões de sobreposição.

O procedimento consiste, primeiramente, encontrar os K vizinhos mais próximos de cada amostra x_i. Então, se a maioria desses K vizinhos pertencerem a uma classe diferente da classe da amostra x_i, podemos considerar que x_i está em um região de sobreposição. O valor de $D3$ para uma determinada classe w_i é obtido pela relação dos elementos na região de sobreposição frente ao total de instâncias pertencentes à classe.

Observa-se que quanto mais alto o valor de $D3$ para uma determinada classe, menor o número de amostras de tal classe na região de sobreposição.

3.8 Considerações finais

Este capítulo teve como finalidade apresentar as principais medidas de complexidade presentes na literatura, detalhar seus conceitos e mostrar como elas podem ser calculadas.

Foram apresentados alguns grupos de medidas como: os que avaliam o quão discriminativo são os atributos; os que procuram avaliar até que ponto duas classes são separáveis linearmente; os que tentam encontrar a região de fronteira de decisão entre as classes; os que apresentam um indicativo da dispersão dos dados; aqueles que mensuram o desbalanceamento entre as classes; os que consideram a correlação dos valores dos atributos com as saídas e aqueles que determinam a densidade das classes.

A pesquisa acerca das medidas idealizou construir uma compreensão mais profunda das métricas de complexidade, de forma a embasar o método especificado no próximo capítulo. Nele, é apresentado um sistema de múltiplos classificadores que apoia-se nos conceitos apresentados nesse capítulo para realizar a seleção dinâmica de classificadores em um sistema de classificação.

4 Metodologia

A estratégia proposta consiste em selecionar dinamicamente um ou mais classificadores com base nas características da instância de teste. Uma das principais preocupações é o uso das informações de complexidade dos dados e a precisão dos classificadores pertencentes ao subconjunto de instâncias K que são vizinhas da amostra de teste, obtidas por estimativas em um conjunto de validação. Este processo é semelhante às técnicas OLA (WOODS; KEGELMEYER; BOWYER, 1997) e KNORAs (KO; SABOURIN; BRITTO JR., 2008), exceto que o critério de definição da vizinhança é diferente, levando em consideração também a complexidade dos dados na nova abordagem.

4.1 Conjunto de dados

Antes de detalhar as etapas particulares da presente abordagem, vale mencionar a divisão do conjunto de dados original em três subconjuntos: treino, validação e teste. Para isso, é aplicado o método *hold-out*, que divide aleatoriamente o conjunto de dados preservando a razão entre as classes (KOHAVI, 1995; YADAV; SHUKLA, 2016).

Esses conjuntos são usados da seguinte forma:

- Conjunto de Treino (X_{train}): para obter subconjuntos nos quais os modelos de classificação são treinados;
- Conjunto de Validação (X_{val}): o objetivo é calibrar os modelos, bem como extrair meta-informações, como os vizinhos de um determinado espaço de características dado;
- Conjunto de Teste (X_{test}): avaliação das estratégias de classificação.

Durante os experimentos foi usado um total de 21 problemas de duas classes, consistindo apenas de atributos numéricos e sem valores ausentes. Tais bases são comumente utilizadas na literatura para medir o desempenho de métodos de seleção dinâmica.

Do grupo avaliado, dez conjuntos de dados vêm do *UCI Machine Learning Repository* (DUA; GRAFF, 2017), três do *KEEL Knowledge Extraction based on Evolutionary Learning* (ALCALA-FDEZ et al., 2010), três da *LKC Ludimila Kuncheva Collection* (KUNCHEVA, 2004), dois do *STATLOG* (KING; FENG; SUTHERLAND, 1995), um da *ELENA (Enhanced Learning for Evolutive Neural Architeture)* (AVILES-CRUZ et al., 1995) e duas bases artificiais geradas com a *toolbox Matlab PRTOOLS* (DUIN et al., 2004). A Tabela 3 apresenta mais detalhes sobre cada conjunto de dados.

Tabela 3 – Principais características das bases usadas nos experimentos

Base	Instâncias	Treino	Teste	Validação	Atributos	Classes	Fonte
Adult	690	345	172	173	14	2	UCI
Banana	2000	1000	500	500	2	2	PRTools
Blood	748	374	187	187	4	2	UCI
Diabetes	766	383	192	191	8	2	UCI
German	1000	500	250	250	24	2	STATLOG
Haberman	306	153	76	77	3	2	UCI
Heart	270	135	67	68	13	2	STATLOG
ILPD	583	292	145	146	10	2	UCI
Ionosphere	350	176	87	87	34	2	UCI
Laryngeal1	213	107	53	53	16	2	LKC
Lithuanian	2000	1000	500	500	2	2	PRTools
Liver	345	173	86	86	6	2	UCI
Magic	19020	9510	4755	4755	10	2	KEEL
Mammo	830	415	207	208	5	2	KEEL
Monk	432	216	108	108	6	2	KEEL
Phoneme	5404	2702	1351	1351	5	2	ELENA
Sonar	208	104	52	52	60	2	UCI
Thyroid	692	346	173	173	16	2	LKC
Vertebral	300	150	75	75	6	2	UCI
WBC	569	285	142	142	30	2	UCI
Weaning	302	151	75	76	17	2	LKC

Considerando esse cenário, a abordagem se apresenta como um processo envolvendo cinco etapas {A, B, C, D e E} conforme representado na Figura 20. A primeira consiste na criação de vários classificadores. As próximas etapas são utilizadas para avaliar os modelos quanto à complexidade dos dados referentes às instâncias de validação e teste. Por fim, a seleção dos classificadores mais adequados no pool para classificar cada instância de teste por vez. Os detalhes relativos a cada etapa são descritos a seguir.

4.2 Geração do conjunto de classificadores base (A)

Na fase de treinamento (Figura 20-A), um pool C composto de M subconjuntos extraídos do conjunto treino X_{train} é criado com base no método *Bagging* (BREIMAN, 1996). Nesse processo, instâncias são selecionadas aleatoriamente com reposição para compor cada um dos subconjuntos M. Embora o processo seja aleatório, as proporções originais entre as classes são mantidas. Cada um dos subconjuntos formados é usado para treinar um classificador base.

Os experimentos foram realizados em um cenário com 20 repetições para cada problema. Para cada repetição, os conjuntos de dados foram divididos pelo método *hold-out*, que divide os conjuntos aleatoriamente e preserva as proporções das classes. As distribuições adotadas foram de 50% para o conjunto de treinamento X_{train}, 25% para o conjunto de dados de validação X_{val} e 25% para o conjunto de dados de teste X_{test}.

Um total de 100 subconjuntos foram gerados usando o método *Bagging* considerando

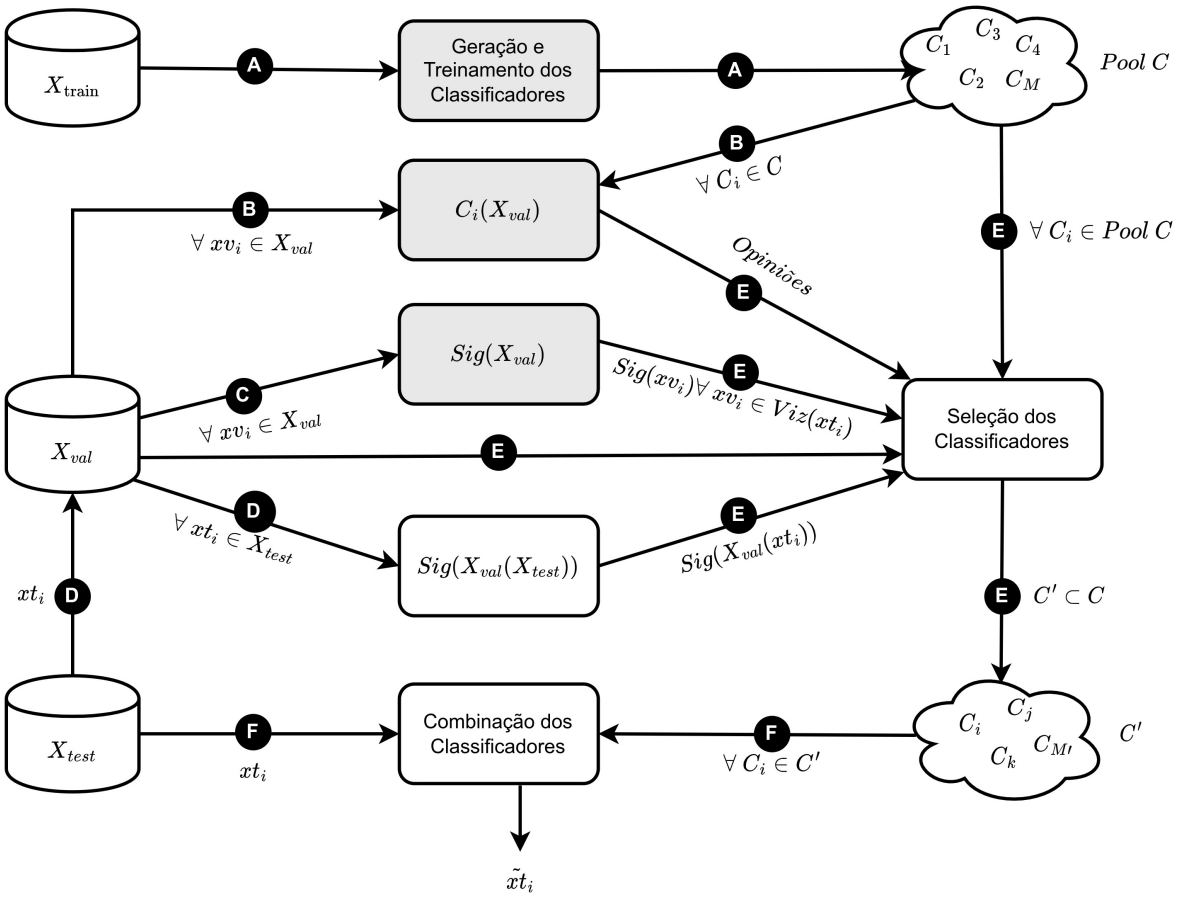


Figura 20 – Visão geral da estratégia

o conjunto de dados de treinamento X_{train} para treinar 100 perceptrons e formar o *pool* de classificadores C. Cada subconjunto construído possui 20% do tamanho do conjunto de treinamento X_{train}. Para estimar a assinatura de complexidade de cada instância de teste e validação, foi escolhida uma vizinhança de tamanho $K = 30$. Este valor mostrou-se apropriado em estudos anteriores (BRUN et al., 2016; BRUN et al., 2018). Na fase de seleção do classificador dinâmico, uma vizinhança de tamanho $k = 7$ foi usado. Este valor tem sido usado em vários trabalhos na literatura para essas estratégias (KO; SABOURIN; BRITTO JR., 2008; CRUZ; CAVALCANTI; REN, 2011; BRUN et al., 2016; CRUZ et al., 2015; BRUN et al., 2018; CRUZ; SABOURIN; CAVALCANTI, 2018).

4.3 Opinião do conjunto de classificadores base para validação (B)

Nesta etapa (Figura 20-B), os classificadores pertencentes ao pool ($\forall\ C_i \in C$), formados na etapa anterior, opinam sobre cada instância do conjunto de validação X_{val}. Portanto, a opinião de cada classificador é agora conhecida e também se ele classifica corretamente ou incorretamente cada padrão do conjunto X_{val}.

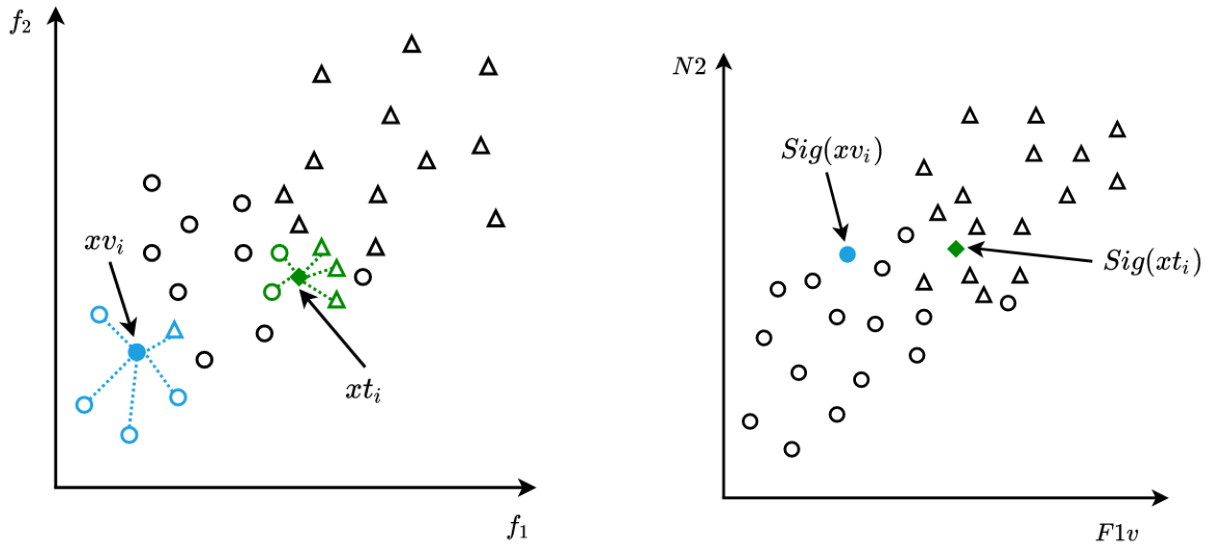
A opinião dos classificadores é um dado importante. Ajuda a avaliar a competência

dos membros do ensemble como parte da etapa de seleção. Os modelos mais adequados são aqueles que apresentam mais previsões corretas na vizinhança do padrão.

4.4 Assinatura de complexidade de instâncias de validação (C)

O objetivo desta etapa é obter a assinatura de complexidade $Sig(xv_i)$ de cada instância xv_i tal que $\forall\ xv_i \in X_{val}$, onde $Sig(xv_i)$ é um vetor de complexidades calculadas. A estimativa do espaço de complexidade, com base no conceito de espaço de características, requer medidas de complexidade definidas a priori. Assim, é possível determinar cada posição xv_i com base na assinatura de complexidade $Sig(xv_i)$ (Figura 20-C).

Todas as vizinhanças de xv_i são estimadas em X_{val}, usando o algoritmo do vizinho mais próximo e a distância euclidiana como critérios. A Figura 21a mostra o espaço de características (composto pelos atributos f_1 e f_2) de X_{val}, onde os K vizinhos de uma instância xv_i (círculo azul) e uma determinada instância formam um subconjunto cuja assinatura de complexidade é calculada. Neste caso particular, a assinatura é composta apenas por medidas $N2$ e $F1v$. Este processo é repetido para todas as instâncias pertencentes a X_{val}, criando um novo conjunto de dados, onde cada métrica de complexidade é um atributo deste novo espaço, conforme mostrado na Figura 21b.



(a) Espaço de características (b) Espaço de complexidade

Figura 21 – Espaço de características e representação do espaço de complexidade

4.5 Assinatura de complexidade de instâncias de teste (D)

Uma vez coletadas as assinaturas de complexidade das instâncias de validação, as assinaturas das instâncias de teste $Sig(xt_i)$ são estimadas para todas as $xt_i \in X_{test}$, respeitando o mesmo espaço de complexidade de xv_i (Figura 20-D).

A assinatura de complexidade de teste $Sig(xt_i)$ considera os vizinhos da instância de teste xt_i de acordo com a estimativa de xv_i em X_{val}. A Figura 21a mostra o espaço de características do conjunto de dados de validação X_{val}. Uma instância de teste xt_i (losango verde) é projetada e a identificação de K (na ilustração adotou-se um $K = 5$) vizinhos de xt_i é realizada. Esses vizinhos formam um subconjunto que é usado para obter a assinatura de complexidade no espaço $F1v$ x $N2$. Em outra etapa, a Figura 21b mostra o espaço de complexidade, onde estão presentes as assinaturas de complexidade de todas as instâncias presentes em X_{val}. Além disso, destaca-se a assinatura $Sig(xt_i)$, que é a projeção da instância de teste xt_i no mesmo espaço de complexidade de xv_i.

4.6 Seleção de classificadores (E)

Nesta fase (Figura 20-E), a preocupação é selecionar um ou mais classificadores que se espera que classifiquem corretamente uma determinada instância de teste. As técnicas de seleção dinâmica utilizam o desempenho dos classificadores na vizinhança do teste, atendendo a critérios como acurácia. A seleção dinâmica aqui apresentada é baseada no espaço de complexidade, então a ideia é combinar pontuações de medida de complexidade com o espaço de características usado para treinar inicialmente os classificadores. Assumindo o uso de todas as medidas da *Data Complexity Library (DCoL)* de Orriols-Puig, Macià e Ho (2010), 14 medidas podem ser adicionadas ao vetor de características originalmente presentes no conjunto de dados. No entanto, o número de medidas de complexidade pode variar em relação a alguns critérios para escolha daquelas consideradas mais adequadas. Um critério pode ser, por exemplo, o *score* de ganho de informação desses novos atributos (*score* de medidas de complexidade).

Os métodos cujas referências estão listadas na Tabela 4 foram executados a partir do conjunto de dados original e sobre o conjunto de dados original combinado com os recursos de complexidade. Seus desempenhos foram, então, comparados.

Tabela 4 – Métodos de Seleção Dinâmica (DS) incluídos na análise comparativa

Métodos de Seleção Dinâmica	Referências
Acurácia Local Total (OLA)	(WOODS; KEGELMEYER; BOWYER, 1997)
Acurácia Local da Classe (LCA)	(WOODS; KEGELMEYER; BOWYER, 1997)
KNORA-Eliminate (KNORA-E)	(KO; SABOURIN; BRITTO JR., 2008)
KNORA-Union (KNORA-U)	(KO; SABOURIN; BRITTO JR., 2008)

Para a estratégia OLA, é escolhido o classificador com maior precisão na região ao redor do teste (projetado no conjunto de validação). Para o LCA, é selecionado o classificador que atingir o maior número de vizinhos cuja classe seja semelhante àquela atribuída ao teste. Para o método KNORAU, todos os classificadores que acertam pelo menos um vizinho da instância de teste são selecionados. No KNORAE, por outro lado,

apenas os classificadores que acertam a todos os vizinhos do padrão de teste são escolhidos. Se nenhum classificador o fizer, o número de vizinhos é reduzido até que pelo menos um dos classificadores possa satisfazer esse requisito.

Em cenários em que há empate entre os classificadores presentes no pool, o método de seleção escolherá todos aqueles que demonstraram a maior competência na vizinhança da instância de teste. Quando mais que um classificador for escolhido, faz-se necessária uma última etapa no processo de classificação (Figura 20-F): a combinação. Neste processo foi adotado a estratégia do voto majoritário, visto que é a abordagem proposta pelos autores das quatro estratégias avaliadas no estudo.

4.7 Avaliando as medidas de complexidade

Dependendo da complexidade do problema a ser analisado, um estimador de complexidade pode enfatizar o quão difícil é a tarefa de classificar. Medidas de complexidade são descritores nos quais várias características do conjunto de dados podem ser observadas.

Com a finalidade de trabalhar com as medidas que mais contribuem para identificar a complexidade dos problemas, é realizada uma análise de ganho de informação para cada uma das 14 medidas estimadas pela *Data Complexity Library (DCoL)* Orriols-Puig, Macià e Ho (2010). A análise foi baseada nas Equações 4.1 e 4.2. O termo p_i corresponde à probabilidade de ocorrência de cada classe i pertencente ao conjunto S. Na segunda equação, do ganho de informação, cada atributo A tem seu ganho calculado. No nosso cenário, tais atributos são as medidas de complexidade. O termo v refere-se a todos os valores que o atributo A pode assumir.

$$Entropia(S) = -\sum_{i=1}^{n} p_i log_2 p_i \tag{4.1}$$

$$Ganho(S, A) = Entropia(S) - \sum_{v \in valores(A)} \frac{|S_v|}{|S|} Entropia(S_v) \tag{4.2}$$

O cálculo, baseado na análise de entropia, foi realizado em 20 repetições para cada um dos 21 problemas apresentados no estudo. O objetivo foi identificar o ganho de informação alcançado a partir de cada medida. Quanto o maior o valor do ganho, mais a medida contribui para o processo de classificação. Para tanto, ranqueamos as medidas de acordo com o valor do ganho para cada problema do estudo. O melhor rank (1ł posição) é atribuído à medida com maior ganho de ganho de informação. Posteriormente calculamos o ranking média de todas os descritores de complexidade. A tabela 5 mostra as classificações médias de ganho de informação obtidas nesta análise.

Tabela 5 – Classificação média de ganho de informação para as medidas de complexidade

Métrica	Classificação média
F1v	1,10
F1	2,29
N2	3,48
L1	3,86
F2	5,90
F3	7,38
L2	7,38
N1	7,57
N4	8,48
N3	9,43
F4	10,67
T1	11,62
L3	11,86
T2	14,00

Os valores apresentados mostram algumas medidas com ganho de informação considerável, como $F1$, $F1v$, $N2$ e $L1$. Por outro lado, existem medidas em que o ganho de informação obtido é mínimo, como $F4$, $T1$, $L3$ e $T2$.

Após a análise do ganho de informação, apenas as medidas $F1$, $F1v$, $N2$ e $L1$ foram selecionadas por enfatizarem a complexidade dos 21 problemas de classificação considerados. Vale ressaltar que a inclusão de outras medidas nos experimentos, geralmente, degradava o desempenho da estratégia de seleção proposta.

As quatro métricas de complexidade selecionadas foram adicionadas ao conjunto original de atributos, criando um novo espaço de representação que combina as características originais das instâncias com seus descritores de complexidade. Este processo é ilustrado na Figura 22. As características originais (f_1 e f_2) foram agrupadas às medidas de complexidade, neste caso, representadas somente por $N2$ por simplicidade. Assim, as características do conjunto de dados original têm sua dimensionalidade aumentada em quatro unidades ($F1$, $F1v$, $L1$ e $N2$).

O processo de adição dos novos quatro atributos é realizado para todas as instâncias de X_{val} e todos os elementos de X_{test}, mantendo a consistência dos atributos nas etapas de treinamento e teste. Os classificadores selecionados na nova representação de dados, de acordo com os k vizinhos mais próximos de cada instância de teste, são definidos no conjunto de validação atualizado, levando em consideração as características dos dois espaços de representação.

A decisão de agrupar os dados originais com os quatro descritores de complexidade deu-se a partir da análise de desempenho da adoção exclusiva das medidas de complexidade como critério de seleção. Os experimentos que levaram em conta apenas os descritores

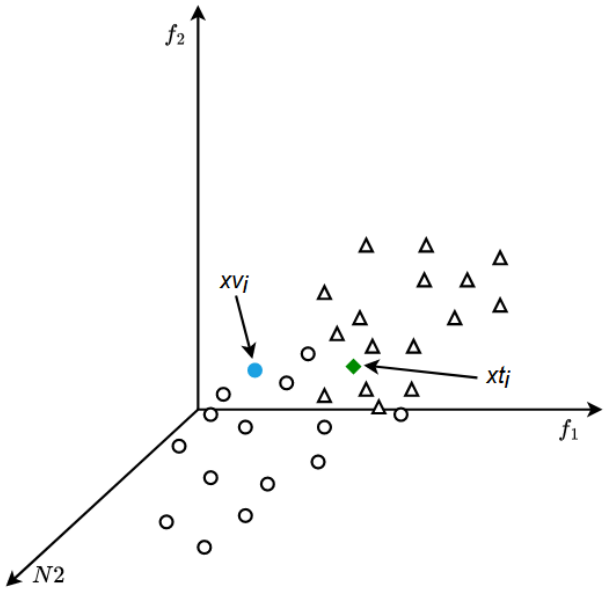


Figura 22 – Representação do novo espaço de característica, combinando a característica original com as medidas de complexidade.

$F1$, $F1v$, $L1$ e $N2$ como critério de vizinhança apresentaram resultados inferiores em comparação à abordagem que combinou os dois conjuntos de características.

4.8 Considerações Finais

Neste capítulo, foi descrita a estratégia do proposta, que utiliza as medidas de complexidade e a acurácia como critérios de seleção dos classificadores mais adequados para cada exemplar de teste. Nele, foram detalhados a estrutura da estratégia proposta e o funcionamento de cada uma de suas etapas. No próximo capítulo, serão apresentados os parâmetros, decisões de projeto e recursos utilizados nos procedimentos adotados durante a execução dos experimentos bem como as percepções alcançadas a partir destes.

5 Experimento e Resultados

No capítulo anterior foi apresentada, de forma generalista, a estratégia proposta. Neste capítulo são apresentados os resultados alcançados com base nas configurações de parâmetros apresentados no capítulo anterior.

Os resultados obtidos nas vinte execuções são mostrados na Tabela 6. Os valores correspondem à acurácia média das 20 repetições do método para cada um dos problemas estudados. Além do valor médio da precisão, o desvio padrão também é fornecido.

A Tabela 6 mostra o desempenho das quatro estratégias de seleção dinâmica usando o conjunto de dados original: OLA, LCA, KNORAU e KNORAE. Além disso, o desempenho das quatro estratégias é mostrado usando os conjuntos de dados combinados (características e descritores de complexidade), aqui referidos como OLA_C, LCA_C, KNORAU_C e KNORAE_C. Então, cada estratégia é comparada com a variação no conjunto base, ou seja, OLA vs. OLA_C e assim sucessivamente. Os valores em negrito correspondem ao melhor desempenho para aquela base/método.

Considerando os cenários estudados, foram feitas 84 comparações, 21 para cada variante do método de seleção. Para ilustrar os comportamentos observados, a Figura 23 apresenta o número de casos observados para cada caso. O número de vezes em que adotar as medidas de complexidade juntamente com as características originais levou a ganha de acurácia está representado nas colunas em verde. Os casos em que a seleção apresentou o melhor desempenho ao se adotar apenas os atributos originais das bases são representados nas colunas em azul. Por fim, nas colunas na cor laranja estão representadas as quantidades de vezes em que não se observou variação na acurácia do método ao se adotar as medidas de complexidade ou o conjunto de dados original.

A comparação do desempenho do OLA é mostrada no primeiro grupo da Figura 23 (OLA vs OLA_C). Nesse cenário, observou-se que as medidas de complexidade proporcionaram um aumento na acurácia média em 10 ocasiões. Em sete casos, houve diminuição da acurácia quando tais medidas foram aplicadas e, nos quatro cenários restantes, não houve alteração no desempenho médio da estratégia.

A análise do desempenho do método LCA (LCA vs. LCA_C) mostrou que: em dez das 21 comparações, houve melhora na acurácia média do método; em quatro ocasiões, houve diminuição na taxa de acerto; e, nas sete bases restantes, não houve alteração na precisão da estratégia de seleção.

O método de seleção dinâmica KNORA-Union (KU vs. KU_C) foi o menos suscetível à variação no conjunto de atributos utilizados na seleção do classificador. Foram

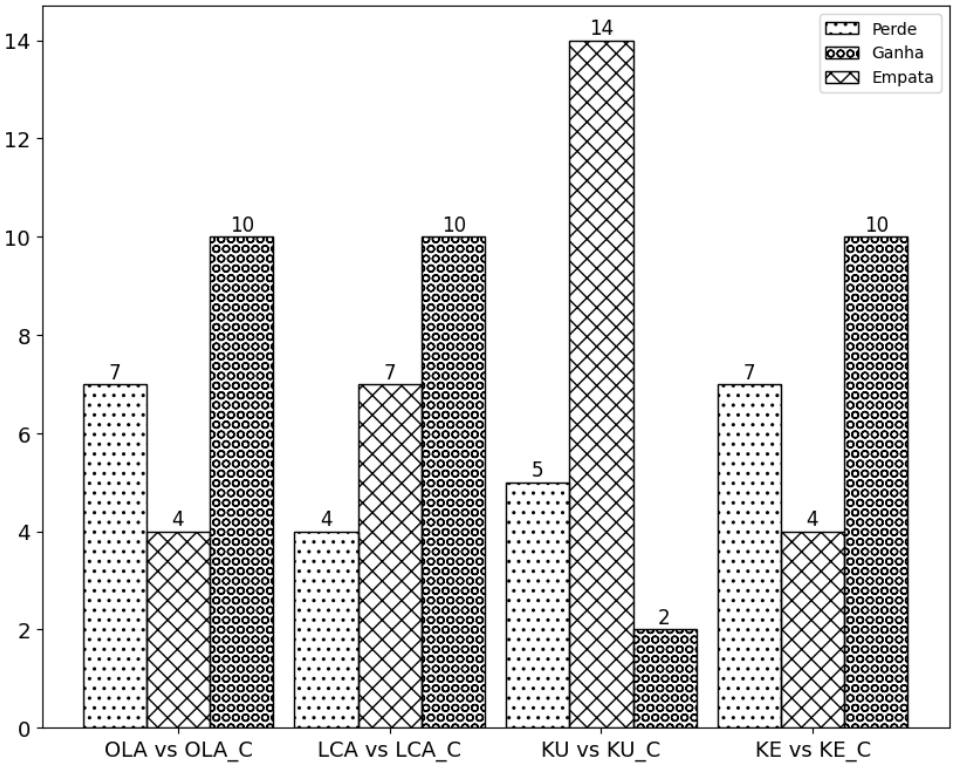


Figura 23 – Gráfico comparativo entre os desempenhos das abordagens com e sem a adição das medidas de complexidade. Em verde as vitórias da técnica que adota as medidas, em azul as derrotas e em laranja os empates

observadas 14 ocasiões onde não houve variação no desempenho médio do método. Além disso, o uso de medidas de complexidade aumentou a precisão do algoritmo em dois casos, sendo que, nas cinco bases restantes, foi mais interessante utilizar apenas os atributos originais das bases.

O comportamento da estratégia de seleção KNORA-Eliminate (KE vs. KE_C) foi semelhante ao OLA. A adoção de medidas de complexidade foi mais precisa em dez casos. Em sete cenários, essa adição levou a uma diminuição na precisão e, nos quatro casos restantes, não houve alteração no desempenho do KNORA-E.

Variações no desempenho das abordagens comparadas aconteceram conforme observado no desvio padrão na Tabela 6. Analisando as repetições, o comportamento do novo método, em uma comparação justa com os demais, mostra que, quando o ganho de informação é maior, a acurácia comumente é maior, sugerindo, uma melhoria no processo de seleção de classificadores. Da mesma forma, o oposto é provado ser verdade.

Tabela 6 – Comparação das performances médias nas 20 execuções. Os desvios padrão correspondentes são dados entre parênteses.

Dataset	OLA	OLA_C	LCA	LCA_C	KNORAU	KNORAU_C	KNORAE	KNORAE_C
Adult	85.0015 (2.00)	**85.0885 (2.01)**	83.9315 (2.13)	**84.1915 (1,96)**	86,7065 (1.90)	86.7065 (1.90)	85.0015 (2.00)	**85.0885 (2.01)**
Banana	95.1100 (1.12)	**95.1200 (1.14)**	95.4500 (1.17)	**95.4600 (1.19)**	90.8009 (1.19)	**90.9000 (1.19)**	95.1100 (1.12)	**95.1200 (1.14)**
Blood	74.0910 (2.16)	**74.1175 (2.14)**	74.1445 (2.12)	**74.1710 (2.10)**	76.1250 (0.31)	76.1250 (0.31)	74.0910 (2.16)	**74.1175 (2.14)**
Diabetes	69.7395 (3.05)	69.7395 (3.05)	70.9125 (3.30)	70.9125 (3.30)	**66.2505 (2.32)**	66.2245 (2.33)	69.7395 (3.05	69.7395 (3.05)
German	**72.8000 (2.16)**	72.7400 (2.07)	71.1200 (3.13)	**71.2400 (3.12)**	76.7800 (1.78)	76.7800 (1.78)	**72.8000 (2.16)**	72.7400 (2.07)
Haberman	71.1695 (4.63)	**71.2995 (4.67)**	71.7550 (4.63)	**71.8195 (4.88)**	73.5075 (1.76)	73.5075 (1.76)	71.1695 (4.63)	**71.2995 (4.67)**
Heart	**80.5125 (4.04)**	80.4390 (3.54)	79.4105 (4.13)	**79.6310 (3.88)**	83.4550 (4.29)	83.4550 (4.29)	**80.5125 (4.04)**	80.4390 (3.54)
ILPD	68.6290 (3.29)	**68.6990 (3.11)**	69.5195 (2.41)	**69.6915 (2.38)**	71.0270 (3.67)	**71.0950 (3.72)**	68.6290 (3.29)	**68.6990 (3.11)**
Ionosphere	85.6810 (3.86)	**85.6820 (3.66)**	85.1135 (4.71)	**85.5115 (5.05)**	83.6360 (2.54)	83.6360 (2.54)	85.6810 (3.86)	**85.6820 (3.66)**
Laryngeal1	81.5745 (3.03)	**81.8515 (2.79)**	80.4635 (5.32)	**80.6490 (4.86)**	83.7060 (4.75)	83.7060 (4.76)	81.5745 (3.04)	**81.8515 (2.80)**
Lithuanian	96.5500 (0.73)	**96.5600 (0.72)**	96.6900 (0.73)	**96.7000 (0.74)**	**94.1500 (0.94)**	94.0100 (0.97)	96.5500 (0.73)	**96.5600 (0.72)**
Liver	64.5990 (3.90)	64.5990 (3.90)	**65.8635 (3.62)**	65.8060 (3.63)	66.7255 (4.44)	66.7255 (4.44)	64.5990 (3.90)	64.5990 (3.90)
Magic	**79.6500 (0.44)**	79.6490 (0.44)	**79.9125 (0.46)**	79.9120 (0.46)	75.9250 (0.54)	75.9250 (0.54)	**79.6500 (0.44)**	79.6490 (0.44)
Mammo	80.6010 (2.59)	**80.6255 (2.70)**	80.2405 (2.01)	80.2405 (2.21)	**82.0675 (2.71)**	82.0435 (2.74)	80.6010 (2.59)	**80.6255 (2.70)**
Monk	90.5550 (2.22)	**90.6475 (2.55)**	87.8235 (2.39)	**88.2875 (2.13)**	**77.7780 (3.05)**	77.6855 (2.96)	90.5550 (2.22)	**90.6475 (2.55)**
Phoneme	**82.8060 (1.08)**	82.7955 (1.09)	**83.3650 (1.01)**	83.3545 (1.02)	77.1980 (1.35)	77.1980 (1.35)	**82.8060 (1.08)**	82.7955 (1.09)
Sonar	75.2885 (7.28)	**75.8655 (6.38)**	73.6535 (6.37)	**74.5195 (5.88)**	74.8070 (5.33)	**74.9990 (5.55)**	75.2885 (7.28)	**75.8655 (6.38)**
Thyroid	**96.1860 (1.25)**	96.1570 (1.10)	96.5610 (1.37)	**96.6190 (1.26)**	96.7920 (0.89)	96.7920 (0.89)	**96.1860 (1.25)**	96.1570 (1.10)
Vertebral	**82.5330 (2.96)**	82.5303 (3.54)	**83.3335 (3.34)**	83.1325 (3.63)	84.7335 (3.68)	84.7335 (3.68)	82.5330 (2.96)	82.5330 (3.54)
WBC	91.9240 (2.31)	91.9240 (2.31)	92.4125 (2.47)	92.4125 (2.47)	91.1550 (2.03)	91.1550 (2.03)	91.9240 (2.31)	91.9240 (2.31)
Weaning	80.4590 (3.23)	**81.0525 (3.26)**	79.5385 (4.26)	79.5385 (4.26)	81.8420 (3.93)	81.8420 (3.93)	80.4590 (3.23)	**81,0525 (3.26)**

Em uma análise de grupo do desempenho dos métodos, verificou-se que a acurácia média aumentou em 32 casos (39,51% do total) quando as características de complexidade foram usadas em conjunto com as características originais. Ademais, os métodos apresentaram desempenho semelhante em 29 ocasiões (35,8%), independentemente dos atributos utilizados na seleção e, por fim, observou-se uma diminuição na acurácia média dos métodos em 23 cenários (28,4%). Além disso, para os vinte problemas avaliados em 17 ocasiões o melhor desempenho geral foi alcançado levando em conta as medidas de complexidade na seleção dos classificadores.

5.1 Considerações finais

Neste capítulo, foram apresentados os experimentos e a análise realizada com a finalidade de avaliar a estratégia proposta. Foi objeto de estudo o desempenho de quatro métodos clássicos da DS com espaço do problema alterado. A avaliação da proposta mostrou-se positiva nos métodos OLA, LCA e KNORAE, onde houve aumento de desempenho na maioria dos conjuntos de dados testados. Enquanto, no método KNORAU o desempenho foi semelhante, visto que o método é menos criterioso na escolha do classificador.

6 Conclusões

Neste trabalho, propomos uma abordagem de Seleção Dinâmica na qual a complexidade do problema de classificação é considerada como parte do espaço de características. A premissa é de que seria positivo, em termos de acurácia, selecionar classificadores cujo desempenho fosse melhor em vizinhanças cuja complexidade é similar à instância de teste em análise.

Com o objetivo de avaliar a aplicabilidade e contribuição da abordagem proposta realizou-se um experimento composto por 21 problemas binários sob um execução composta de 20 repetições. Para cada problema comparou-se o desempenho de quatro abordagens clássicas de seleção dinâmica de classificadores (OLA, LCA, KNORAU e KNORAE) usando apenas o conjunto original de características frente ao desempenho alcançando pelas mesmas abordagens mas considerando como espaço de características a combinação dos atributos originais com descritores de complexidade.

As medidas provaram apresentar um índice de informação considerável e elevaram a acurácia das abordagens clássicas de DS. Apesar disso, nem todas as medidas contribuem para a precisão do sistema, principalmente aquelas com baixo *score* de ganho de informação. As medidas mais influentes e que foram adotadas nos experimentos foram $F1$, $F1v$, $L1$ e $N2$.

A nova abordagem é promissora. Os resultados mostram que em $71,6\%$ dos casos observou-se um cenário positivo e que em apenas $28,4\%$ dos cenários testados observou-se piora na acurácia. Além disso, em 85% dos problemas o melhor desempenho geral foi alcançado quando os atributos de complexidade foram empregados.

Os experimentos realizados levaram em conta apenas abordagens clássicas de DS, entretanto, a técnica explorada pode ser usada para abordagens mais recentes e também projetar novos métodos que assumem medidas de complexidade, como critérios para selecionar os classificadores mais adequados de acordo com cada problema em particular.

Medidas de complexidade lidam com diferentes aspectos do problema de classificação, portanto, em pesquisas futuras, métodos que explorem essas diferenças como um aspecto de complementaridade devem ser considerados. Este trabalho apresenta o ganho de informação como critério para selecionar medidas de menor complexidade. No entanto, outros critérios também podem ser interessantes.

Outro aspecto da pesquisa a ser observado são as medidas de complexidade. Neste trabalho elas foram escolhidas estaticamente, ou seja, as mesmas medidas são adotadas para todos os conjuntos de dados e repetições. Pesquisas futuras podem considerar escolher

essas medidas dinamicamente, de acordo com cada padrão e conjunto de dados.

Na formação do novo espaço do problema foram combinados todos os atributos originais do conjunto de dados com as medidas de complexidades selecionadas. Pode ser interessante realizar um estudo para avaliar se a utilização de um subconjunto dos atributos originais combinados com as medias de complexidade poderia levar a melhores resultados.

Referências

ALCALA-FDEZ, J. et al. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, v. 17, p. 255–287, 01 2010. Citado na página 61.

AVILES-CRUZ, C. et al. *Databases, Enhanced Learning for Evolutive Neural Architecture*. [S.l.], 1995. 47 p. Disponível em: <http://www.dice.ucl.ac.be/neural-nets/Research/Projects/ELENA/elena.htm>. Citado na página 61.

BELLMAN, R. E. *Adaptive Control Processes*. Princeton University Press, 2015. Disponível em: <https://doi.org/10.1515/9781400874668>. Citado na página 26.

BREIMAN, L. Bagging predictors. *Mach. Learn.*, Kluwer Academic Publishers, USA, v. 24, n. 2, p. 123–140, ago. 1996. ISSN 0885-6125. Disponível em: <https://doi.org/10.1007/BF00058655>. Citado 2 vezes nas páginas 24 e 62.

BRITTO JR., A. S.; SABOURIN, R.; OLIVEIRA, L. E. Dynamic selection of classifiers—a comprehensive review. *Pattern Recognition*, v. 47, n. 11, p. 3665 – 3680, 2014. ISSN 0031-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320314001885>. Citado 7 vezes nas páginas 10, 18, 22, 23, 29, 31 e 42.

BROWN, G. et al. Diversity creation methods: a survey and categorisation. *Information Fusion*, v. 6, n. 1, p. 5–20, 2005. ISSN 1566-2535. Diversity in Multiple Classifier Systems. Citado 2 vezes nas páginas 23 e 28.

BRUN, A. L. et al. Contribution of data complexity features on dynamic classifier selection. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2016. p. 4396–4403. Citado 3 vezes nas páginas 19, 42 e 63.

BRUN, A. L. et al. A framework for dynamic classifier selection oriented by the classification problem difficulty. *Pattern Recognition*, v. 76, p. 175–190, 2018. ISSN 0031-3203. Citado 4 vezes nas páginas 19, 27, 42 e 63.

CRUZ, R.; CAVALCANTI, G.; REN, T. I. A method for dynamic ensemble selection based on a filter and an adaptive distance to improve the quality of the regions of competence. In: *Proceedings of the International Joint Conference on Neural Networks*. [S.l.: s.n.], 2011. p. 1126 – 1133. Citado na página 63.

CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Meta-des.oracle: Meta-learning and feature selection for dynamic ensemble selection. *Information Fusion*, v. 38, p. 84–103, 2017. ISSN 1566-2535. Disponível em: <https://www.sciencedirect.com/science/article/pii/S156625351730101X>. Citado na página 19.

CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, v. 41, p. 195 – 216, 2018. ISSN 1566-2535. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1566253517304074>. Citado 8 vezes nas páginas 10, 18, 22, 23, 29, 30, 31 e 63.

CRUZ, R. M. et al. Meta-des: A dynamic ensemble selection framework using meta-learning. *Pattern Recognition*, v. 48, n. 5, p. 1925–1935, 2015. ISSN 0031-3203. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0031320314004919>. Citado 3 vezes nas páginas 19, 31 e 63.

DIDACI, L. et al. A study on the performances of dynamic classifier selection based on local accuracy estimation. *Pattern Recognition*, v. 38, n. 11, p. 2188–2191, 2005. ISSN 0031-3203. Citado 3 vezes nas páginas 19, 33 e 34.

DUA, D.; GRAFF, C. *UCI Machine Learning Repository*. 2017. Disponível em: <http://archive.ics.uci.edu/ml>. Citado na página 61.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2. ed. New York: Wiley, 2001. ISBN 978-0-471-05669-0. Citado 2 vezes nas páginas 17 e 22.

DUIN, R. et al. *Prtools, a matlab toolbox for pattern recognition*. 2004. Citado na página 61.

FREUND, Y.; SCHAPIRE, R. E. Experiments with a new boosting algorithm. In: *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1996. (ICML'96), p. 148–156. ISBN 1558604197. Citado 2 vezes nas páginas 24 e 25.

GIACINTO, G.; ROLI, F. Methods for dynamic classifier selection. In: *Proceedings 10th International Conference on Image Analysis and Processing*. [S.l.: s.n.], 1999. p. 659–664. Citado 3 vezes nas páginas 19, 33 e 34.

GIACINTO, G.; ROLI, F. Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recognition*, v. 34, n. 9, p. 1879–1881, 2001. ISSN 0031-3203. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0031320300001503>. Citado na página 19.

GUNES, V. et al. Combination, cooperation and selection of classifiers: A state of the art. *International Journal of Pattern Recognition and Artificial Intelligence*, v. 17, n. 08, p. 1303–1324, 2003. Disponível em: <https://doi.org/10.1142/S0218001403002897>. Citado 6 vezes nas páginas 18, 22, 29, 38, 39 e 40.

HO, T. K. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 8, p. 832–844, 1998. Citado 3 vezes nas páginas 24, 25 e 26.

HO, T. K.; BASU, M. Complexity measures of supervised classification problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 3, p. 289–300, 2002. Citado 6 vezes nas páginas 18, 19, 43, 44, 54 e 55.

KING, R. D.; FENG, C.; SUTHERLAND, A. Statlog: Comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence*, 1995. Citado na página 61.

KITTLER, J. Combining classifiers: A theoretical framework. *Pattern Analysis and Applications*, v. 1, n. 1, p. 18–27, 1998. ISSN 14337541. Citado 5 vezes nas páginas 18, 22, 38, 39 e 40.

KO, A. H.; SABOURIN, R.; BRITTO JR., A. S. From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, v. 41, n. 5, p. 1718 – 1731, 2008. ISSN 0031-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320307004499>. Citado 10 vezes nas páginas 18, 19, 22, 29, 30, 31, 34, 61, 63 e 65.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (IJCAI'95), p. 1137–1143. ISBN 1558603638. Citado na página 61.

KUNCHEVA, L. Clustering-and-selection model for classifier combination. In: *KES'2000. Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies. Proceedings (Cat. No.00TH8516)*. [S.l.: s.n.], 2000. v. 1, p. 185–188 vol.1. Citado na página 31.

KUNCHEVA, L. A theoretical study on six classifier fusion strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 2, p. 281–286, 2002. Citado na página 31.

KUNCHEVA, L. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley, 2014. ISBN 9781118914540. Disponível em: <https://books.google.com.br/books?id=RtRLBAAAQBAJ>. Citado 6 vezes nas páginas 10, 17, 18, 22, 38 e 40.

KUNCHEVA, L.; WHITAKER, C. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, v. 51, p. 181–207, 05 2003. Citado 3 vezes nas páginas 23, 28 e 36.

KUNCHEVA, L. I. *LKC Ludmila Kuncheva Collection*. 2004. Citado na página 61.

KUNCHEVA, L. I.; RODRIGUEZ, J. J. Classifier ensembles with a random linear oracle. *IEEE Transactions on Knowledge and Data Engineering*, v. 19, n. 4, p. 500–508, 2007. Citado na página 19.

KUNCHEVA, L. I. et al. Complexity of data subsets generated by the random subspace method: An experimental investigation. In: KITTLER, J.; ROLI, F. (Ed.). *Multiple Classifier Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. p. 349–358. ISBN 978-3-540-48219-2. Citado na página 26.

LORENA, A. et al. *Inteligência Artificial: uma abordagem de Aprendizado de Máquina (2a edição)*. [S.l.: s.n.], 2021. ISBN 9788521637493. Citado na página 17.

LORENA, A. et al. Data complexity meta-features for regression problems. *Machine Learning*, v. 107, p. 209–246, 01 2018. Citado 5 vezes nas páginas 18, 44, 57, 58 e 59.

LORENA, A. C. et al. Analysis of complexity indices for classification problems: Cancer gene expression data. *Neurocomputing*, v. 75, n. 1, p. 33–42, 2012. ISSN 0925-2312. Brazilian Symposium on Neural Networks (SBRN 2010) International Conference on Hybrid Artificial Intelligence Systems (HAIS 2010). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0925231211004085>. Citado na página 55.

LORENA, A. C. et al. How complex is your classification problem? a survey on measuring classification complexity. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 52, n. 5, set. 2019. ISSN 0360-0300. Disponível em: <https://doi.org/10.1145/3347711>. Citado 7 vezes nas páginas 18, 44, 45, 48, 49, 54 e 55.

MACIà, N.; ORRIOLS-PUIG, A.; BERNADó-MANSILLA, E. In search of targeted-complexity problems. In: *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*. New York, NY, USA: Association for Computing Machinery, 2010. (GECCO '10), p. 1055–1062. ISBN 9781450300728. Disponível em: <https://doi.org/10.1145/1830483.1830674>. Citado 4 vezes nas páginas 18, 19, 26 e 43.

MALINA, W. Two-parameter fisher criterion. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, v. 31, n. 4, p. 629–636, 2001. Citado na página 46.

MOLLINEDA, R. A.; SáNCHEZ, J. S.; SOTOCA, J. M. Data characterization for effective prototype selection. *In 2nd Iberian Conference on Pattern Recognition and Image Analysis(IbPRIA)*, p. 27–34, 2005. Citado na página 45.

MONTEIRO, M. et al. Classifier pool generation based on a two-level diversity approach. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. [S.l.: s.n.], 2021. p. 2414–2421. Citado na página 27.

OKUN, O.; VALENTINI, G. Dataset complexity can help to generate accurate ensembles of k-nearest neighbors. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, p. 450–457, 2008. Citado na página 42.

ORRIOLS-PUIG, A.; MACIà, N.; HO, T. Dcol: Data complexity library in c++ (documentation). In: . Technical report, La Salle - Universitat Ramon Llull, 2010. Disponível em: <http://dcol.sourceforge.net/>. Citado 4 vezes nas páginas 46, 48, 65 e 66.

PONTI JR., M. P. Combining classifiers: From the creation of ensembles to the decision fusion. In: *2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials*. [S.l.: s.n.], 2011. p. 1–10. Citado 3 vezes nas páginas 26, 28 e 37.

RUTA, D.; GABRYS, B. An overview of classifier fusion methods. *Computing and Information Systems*, v. 7, p. 1–10, 01 2000. Citado 3 vezes nas páginas 38, 41 e 42.

SABOURIN, M. et al. Classifier combination for hand-printed digit recognition. In: *Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR '93)*. [S.l.: s.n.], 1993. p. 163–166. Citado 2 vezes nas páginas 19 e 32.

SÁNCHEZ, J. S.; MOLLINEDA, R. A.; SOTOCA, J. M. An analysis of how training data complexity affects the nearest neighbor classifiers. *Pattern Analysis and Applications*, v. 10, p. 189–201, 2007. Citado 3 vezes nas páginas 42, 59 e 60.

SANTANA, A. et al. A dynamic classifier selection method to build ensembles using accuracy and diversity. In: *2006 Ninth Brazilian Symposium on Neural Networks (SBRN'06)*. [S.l.: s.n.], 2006. p. 36–41. Citado 2 vezes nas páginas 19 e 35.

SANTOS, E. M. dos; SABOURIN, R.; MAUPIN, P. Ambiguity-guided dynamic selection of ensemble of classifiers. In: *2007 10th International Conference on Information Fusion*. [S.l.: s.n.], 2007. p. 1–8. Citado na página 19.

SHIPP, C. A.; KUNCHEVA, L. I. Relationships between combination methods and measures of diversity in combining classifiers. *Information Fusion*, v. 3, n. 2, p. 135–148, 2002. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1566253502000519>. Citado 2 vezes nas páginas 23 e 28.

SOARES, R. et al. Using accuracy and diversity to select classifiers to build ensembles. In: *The 2006 IEEE International Joint Conference on Neural Network Proceedings*. [S.l.: s.n.], 2006. p. 1310–1316. Citado na página 31.

SOTOCA, J.; SáNCHEZ, J.; MOLLINEDA, R. A review of data complexity measures and their applicability to pattern classification problems. *Actas del III Taller Nacional de Mineria de Datos y Aprendizaje*, 01 2005. Citado 3 vezes nas páginas 18, 44 e 59.

SPEARMAN, C. The proof and measurement of association between two things. *The American Journal of Psychology*, University of Illinois Press, v. 15, n. 1, p. 72–101, 1904. ISSN 00029556. Disponível em: <http://www.jstor.org/stable/1412159>. Citado 2 vezes nas páginas 57 e 58.

TANWANI, A. K.; FAROOQ, M. Classification potential vs. classification accuracy: A comprehensive study of evolutionary algorithms with biomedical datasets. In: BACARDIT, J. et al. (Ed.). *Learning Classifier Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 127–144. ISBN 978-3-642-17508-4. Citado na página 56.

WOLOSZYNSKI, T.; KURZYNSKI, M. A probabilistic model of classifier competence for dynamic ensemble selection. *Pattern Recognition*, v. 44, n. 10, p. 2656–2668, 2011. ISSN 0031-3203. Semi-Supervised Learning for Visual Content Analysis and Understanding. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0031320311001245>. Citado na página 19.

WOLOSZYNSKI, T. et al. A measure of competence based on random classification for dynamic ensemble selection. *Information Fusion*, v. 13, n. 3, p. 207–213, 2012. ISSN 1566-2535. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1566253511000297>. Citado na página 19.

WOLPERT, D.; MACREADY, W. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, v. 1, n. 1, p. 67–82, 1997. Citado na página 17.

WOODS, K.; KEGELMEYER, W. P.; BOWYER, K. Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, n. 4, p. 405–410, 1997. Citado 6 vezes nas páginas 19, 31, 32, 33, 61 e 65.

XIAO, J.; HE, C. Dynamic classifier ensemble selection based on gmdh. In: *2009 International Joint Conference on Computational Sciences and Optimization*. [S.l.: s.n.], 2009. v. 1, p. 731–734. Citado na página 19.

YADAV, S.; SHUKLA, S. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: *2016 IEEE 6th International Conference on Advanced Computing (IACC)*. [S.l.: s.n.], 2016. p. 78–83. Citado na página 61.

YAN, Y. et al. Sorting-based dynamic classifier ensemble selection. In: *2013 12th International Conference on Document Analysis and Recognition*. [S.l.: s.n.], 2013. p. 673–677. Citado 2 vezes nas páginas 19 e 36.