

3

Resultados e Discussão

Neste capítulo estão apresentados os resultados e a respectiva discussão associada. Apresentaremos na seção 3.1.1 os resultados referentes a procedimentos de classificação (CQV cuja metodologia foi apresentada na seção 2.1; ZRG cuja metodologia foi apresentada na seção 2.2 e QVFD, cuja metodologia foi apresentada na seção 2.3); na seção 3.2 apresentamos os resultados referentes à seleção de observações (correspondentes à aplicação do CQV a identificação de observações com rótulo invertido, metodologia apresentada no final da seção 2.1 e a aplicação da ZRG na seleção de observações, metodologia mostrada na subseção 2.2.1).

3.1.

Resultados Referentes à Classificação

Nesta seção estão apresentados os resultados e discussões de experimentos de classificação referentes ao desenvolvimento metodológico apresentado nas seções 2.1, 2.2 e 2.3.

3.1.1.

Classificador por Quantização Vetorial (CQV) e Quantizador Vetorial das Fronteiras de Decisão (QVFD)

Nesta subseção apresentam-se os resultados correspondentes aos métodos CQV, descrito na seção 2.1 e QVFD, descrito na seção 2.3.

Em relação ao CQV, para cada experimento, o resultado apresentado é o melhor dentre variações de número de protótipos nos bancos (potências de 2 menores ou iguais a décima parte da quantidade de observações, ou seja, $2^i \leq \#X/10$, $i \in \mathbb{N}$). Já o algoritmo QVFD acrescenta protótipos iterativamente ao longo da fase dentro-da-amostra. Dessa forma, a complexidade do modelo não é determinada pelo projetista, como acontece no CQV e no LVQ.

Os seguintes parâmetros foram utilizados para todos os experimentos: $\varsigma = 10^{-1}$ e $\varepsilon = 10^{-2}$ (algoritmo LBG).

Quatro algoritmos de classificação foram usados com a finalidade de comparação com os métodos propostos: (i) *Learning Vector Quantization* 1 (LVQ 1), onde o estabelecimento do número total de protótipos obedeceu ao mesmo critério estipulado ao CQV. No algoritmo CQV, as observações são quantizadas vetorialmente em uma etapa não supervisionada, utilizando o número total de protótipos. No LVQ, cada protótipo está associado a uma das duas classes, metade destes a C_1 e a outra metade a C_2 . Esta mudança ocorreu pela diferença entre os dois métodos, já que o LVQ exige a atribuição da classe do protótipo antes do processo de aprendizado; foram utilizadas 50 épocas e taxa de aprendizado fixa em 0.01; (ii) Uma rede neural (RN) em camadas, treinada com regularização Bayesiana, com dez neurônios iniciais na camada oculta e função de ativação logística em todos os neurônios; (iii) Máquina de Vetor de Suporte (SVM) com função de bases radiais como *kernel* e (iv) k vizinhos mais próximos, com diferentes valores (3, 5, 7 e 9) para k. Os algoritmos para LVQ e K-NN foram diretamente programados em Matlab, enquanto que para RN e SVM foram utilizadas, respectivamente, as *toolboxes* de *Neural Networks* e *Bioinformatics* também do Matlab.

Foram executados 9 experimentos: **Experimento 1:** Classes divididas através da função cosseno, descrito na subseção 2.4.1; **Experimento 2:** Círculos concêntricos, descrito na subseção 2.4.3; **Experimento 3:** Waveform, descrito na subseção 2.4.4; **Experimento 4:** Letter-B, descrito na subseção 2.4.5; **Experimento 5:** Statlog, descrito na subseção 2.4.6; **Experimento 6:** Diagnóstico de doenças cardíacas, descrito na subseção 2.4.7; **Experimento 7:** Classificação de punção aspirativa por agulha fina de nódulo mamário para diagnóstico, descrito na subseção 2.4.8; **Experimento 8:** Ionosfera, descrito na subseção 2.4.9; **Experimento 9:** Câncer de Pulmão, descrito na subseção 2.4.10.

Para os experimentos 1 a 5, utilizamos para o número de protótipos do LVQ o seguinte procedimento: todas as potências de 2 menores ou iguais a décima parte da quantidade de observações, ou seja, $2^i \leq \#X/10$, $i \in \mathbb{N}$. Para cada uma destas configurações, foi considerada a média de 10 rodadas independentes. O desempenho médio fora-da-amostra referente ao melhor desempenho médio

dentro-da-amostra foi tomado como indicador de desempenho. Para RN, foi considerada a média de 10 rodadas, utilizando o método de Regularização Bayesiana [62] com o intuito de diminuir a arbitrariedade na especificação da arquitetura neural. Para SVM, ambas, a largura do kernel e a constante C sofreram variações entre 2^0 , $2^{0.5}$, 2, 2^2 , 2^3 , 2^4 e 2^5 . Foram tomados os valores fora-da-amostra correspondentes ao melhor desempenho dentro-da-amostra. Em relação a k vizinhos mais próximos, variou-se k nos valores 3, 5, 7 e 9, com o mesmo procedimento do SVM para obtenção do valor fora-da-amostra.

Para os experimentos 6 a 8 (com menos observações), usamos validação cruzada (k-fold) com $k = 10$. Para cada banco de dados, uma partição foi criada e, para cada técnica (LVQ, RN, SVM e K-NN) a fase dentro-da-amostra foi efetuada por um conjunto de treinamento formado por 9 de 10 folds da partição. O fold remanescente foi usado para calcular o desempenho fora-da-amostra. Este procedimento foi repetido 10 vezes, em cada um deles com um fold diferente como conjunto de teste. Para LVQ, K-NN e SVM, foram executados vários procedimentos de validação cruzada, referentes à variação do número de protótipos, do número de vizinhos e da combinação dos parâmetros, respectivamente. Para cada algoritmo, foi selecionado o desempenho fora-da-amostra relacionado ao melhor desempenho de validação cruzada dentro-da-amostra. Para o experimento 9 foi realizado um procedimento de leave-one-out [7] com as 27 observações. Neste experimento houve uma exceção em relação ao número de protótipos utilizados para o CQV e o LVQ. Os dois métodos foram testados com 2 e 4 protótipos em virtude da pouca quantidade de observações. Os resultados obtidos por todos os algoritmos para cada experimento estão na Tabela 1. Na Tabela 2 encontra-se o custo computacional médio em segundos para cada um dos métodos.

Tabela 1: Desempenho de classificação fora-da-amostra (Acerto) e desvio padrão (dp) para os algoritmos CQV e QVFD. Parênteses indicam dentro-da-amostra.

	QVFD		CQV		LVQ		K-NN		RN		SVM	
	Acerto	Dp	Acerto	dp	Acerto	dp	Acerto	dp	Acerto	dp	Acerto	dp
Exp 1	95,9% (98,9%)	-	98,2% (99,3%)	-	98,3% (98,9%)	0,3 0,2	99,1% (99,1%)	-	99,4% (99,7%)	0,3 0,5	99,0% (99,2%)	-
Exp 2	99,2% (99,4%)	-	99,4% (99,6%)	-	98,3% (98,5%)	0,1 0,1	99,5% (99,4%)	-	99,7% (100%)	0,1 0,0	99,0% (98,7%)	-
Exp 3	88,0% (93,7%)	-	87,2% (92,3%)	-	90,6% (95,1%)	0,3 0,6	89,8% (88,2%)	-	89,7% (99,6%)	0,5 0,1	67,6% (100%)	-
Exp 4	95,4% (98,9%)	-	98,3% (98,9%)	-	97,3% (99,2%)	0,2 0,1	99,3% (99,3%)	-	98,6% (99,8%)	0,2 0,2	99,4% (100%)	-
Exp 5	97,5% (98,5%)	-	98,8% (98,7%)	-	98,6% (99,0%)	0,2 0,2	99,6% (98,9%)	-	98,8% (99,9%)	0,3 0,1	98,3% (100%)	-
Exp 6	73,4% (90,2%)	6,7 1,2	79,2% (87,2%)	3,9 0,9	78,1% (85,6%)	5,2 0,8	78,9% (79,1%)	2,6 0,8	74,9% (96,9%)	4,2 0,4	75,0% (100%)	5,1 0,0
Exp 7	96,1% (98,7%)	2,4 0,4	97,1% (99,4%)	1,7 0,2	97,1% (98,0%)	2,1 0,2	97,5% (97,4%)	2,3 0,4	94,9% (99,8%)	3,5 0,3	95,9% (100%)	3,1 0,0
Exp 8	89,2% (95,6%)	4,6 2,3	88,6% (95,0%)	3,5 1,4	86,0% (91,3%)	4,6 1,5	83,8% (84,7%)	5,7 1,1	88,0% (100%)	4,2 0,0	89,5% (100%)	8,0 0,0
Exp 9	85,2% (96,0%)	-	85,2% (93,9%)	-	81,5% (96,4%)	-	85,2% (84,6%)	-	85,2% (100%)	-	70,4% (100%)	-

Tabela 2: Tempo computacional médio em segundos.

	QVFD	CQV	LVQ	K-NN	RN	SVM
Exp 1	1,8	518,4	22,2	5,6	6,1	1,0
Exp 2	0,7	688,7	29,5	8,9	2,7	2,4
Exp 3	2,0	792,6	56,2	48,6	76,1	5,8
Exp 4	6,7	19545,1	301,4	30,6	77,7	23,6
Exp 5	3,7	1936	79,8	88,0	87,5	4,7
Exp 6	1,9	75,2	7,0	0,7	6,1	0,3
Exp 7	0,1	66,0	6,6	0,6	5,2	5,5
Exp 8	0,6	15,4	3,2	0,3	14,5	4,3
Exp 9	0,1	0,4	0,3	0,1	44,5	0,01

Note que houve um considerável equilíbrio entre os seis algoritmos. Para cada experimento, encontra-se em negrito o valor correspondente ao melhor desempenho. RN e K-NN obtiveram três melhores desempenhos. Entretanto, dois destes três experimentos, considerando a rede neural, correspondem aos experimentos simulados. Os algoritmos CQV e SVM conseguiram dois melhores desempenhos, enquanto que LVQ e QVFD conseguiram um melhor desempenho cada.

Dois pontos merecem ser discutidos a partir destes resultados: (i) O equilíbrio encontrado entre LVQ, K-NN, RN e SVM pode ser considerado esperado uma vez que estes métodos são reconhecidamente eficientes para classificação. Vale notar que, buscou-se, nas implementações de cada método, melhorar os desempenhos através de variações na complexidade. Ou seja, os resultados para cada um destes 4 métodos é o melhor que pudemos obter em cada caso. Entende-se por variação de complexidade imposta a LVQ e K-NN a variação do número de protótipos e número de vizinhos, respectivamente; a variação de parâmetros do SVM e a utilização de regularização Bayesiana nas redes neurais. (ii) os bons resultados obtidos pelos algoritmos propostos (CQV e QVFD) mostram, empiricamente, que estes algoritmos podem ser competitivos frente a alguns dos mais reconhecidos algoritmos de classificação, mostrando seu potencial e justificando assim suas propostas.

Note que o CQV obteve dois melhores desempenhos (experimentos 6 e 9) e dois segundos melhores desempenhos (experimento 5 – empatado com RN e experimento 7, empatado com LVQ). Além disso, comparando o CQV com cada um dos algoritmos utilizados para comparação, os resultados foram (vitória-empate-derrota): 6-1-2 contra LVQ, 2-1-6 contra K-NN, 3-2-4 contra RN e 6-0-3 contra SVM. Assim, pode-se considerar que o CQV obteve muito bom desempenho frente ao LVQ e SVM, houve aproximadamente equivalência com RN, sendo realmente pior apenas contra K-NN. Em termos de tempo computacional, observa-se que o CQV foi o algoritmo mais custoso em 7 dos 8 experimentos. Isso ocorre em consequência do processo de quantização vetorial executado no início do algoritmo.

Já o QVFD, embora tenha apresentado resultados inferiores ao CQV, obteve um melhor resultado (experimento 9). Além disso, comparando-se com os outros métodos tem-se (vitória-empate-derrota): 3-0-6 contra LVQ; 1-1-7 contra K-NN; 2-1-6 contra RN e 4-0-5 contra SVM. Considerando o custo computacional, pode-se notar que o algoritmo QVFD apresenta um excelente resultado sendo o mais rápido em 6 dos 9 experimentos e sempre apresentando um baixo custo quando comparado aos outros algoritmos.

Algumas observações podem ser feitas a partir dos resultados obtidos. Como muitos experimentos realizados nesta tese foram executados em bancos de dados desbalanceados, vamos fixar a análise neste tópico. No Experimento 9, por

exemplo, tem-se um banco de dados diferenciado com 27 observações e 56 atributos de entrada. Além desta dificuldade, outro ponto importante é que o banco de dados é desbalanceado, com 8 observações de C_1 e 19 de C_2 . Curiosamente, quatro dos seis métodos obtiveram desempenho de 85,2% (o que corresponde a 4 erros em 27). Investigando as observações classificadas erroneamente por todos os métodos chegamos à conclusão de que: (i) três observações de C_1 (a classe minoritária) foram classificadas erroneamente por 4 dos 6 métodos. Os únicos métodos que obtiveram êxito na classificação destas observações foram o QVFD e a RN; (ii) CQV e SVM erraram apenas observações de C_1 ; (iii) LVQ e K-NN erraram, cada um deles, apenas uma observação de C_2 e o restante de C_1 ; (iv) Finalmente, a surpresa ficou por conta da RN e do QVFD, que erraram 4 observações, todas elas de C_2 , acertando todas as observações da classe minoritária C_1 . A conclusão que pode ser tirada é que, a menos de RN e do QVFD, os métodos foram prejudicados pela dificuldade relativa ao desbalanceamento do banco. Entretanto o bom desempenho do QVFD na classe minoritária não implica em um bom desempenho do algoritmo em problemas com bancos de dados desbalanceados. No Experimento 4, por exemplo, tem-se o conjunto fora-da-amostra formado por 366 observações de C_1 e 9234 observações de C_2 , e o QVFD errou 444 observações, 312 delas pertencentes a C_1 e apenas 132 de C_2 . A RN errou em média apenas 18,7% das observações de C_1 neste experimento enquanto o QVFD errou 85,2%. Já no banco de dados ionosfera (experimento 8), que possui 225 de C_1 contra 126 de C_2 , o QVFD errou ao todo 26 observações entre as 126 de C_2 , a RN errou 31 e o CQV errou 32. Para este experimento, estes desempenhos não são comparáveis ao SVM, que obteve o melhor desempenho geral entre todos os algoritmos errando apenas 2 observações da classe minoritária. No Experimento 5 (banco Statlog), o QVFD errou apenas 16 das 461 observações da classe minoritária; o CQV errou 8 e o K-NN que obteve o melhor desempenho geral errou 5 observações da classe minoritária. Uma possível análise futura pode ser feita sobre o desempenho do QVFD e do CQV em bancos desbalanceados.

3.1.2.

Zona de Risco Generalizada

Nesta subseção apresentam-se os resultados correspondentes a ZRG, descrita na seção 2.2.

Quatro algoritmos de classificação foram usados com a finalidade de comparação com os métodos propostos: (i) *Learning Vector Quantization* 1 (LVQ 1), onde o número de protótipos foi igual a maior potência de 2 menor ou igual a raiz quadrada do total de observações; Foram utilizadas 50 épocas e taxa de aprendizado fixa em 0.01; Uma rede neural (RN) em camadas, treinada com regularização Bayesiana, com dez neurônios iniciais na camada oculta e função de ativação logística em todas as unidades; (iii) Máquina de Vetor de Suporte (SVM) com função de bases radiais como *kernel* e (iv) k vizinhos mais próximos, com k igual a 3. Os algoritmos para LVQ e K-NN foram diretamente programados em Matlab, enquanto que para RN e SVM foram utilizadas, respectivamente, as *toolboxes* de *Neural Networks* e *Bioinformatics* também do Matlab.

Uma vez que nesta parte da tese o objetivo não é comparar um método contra o outro, e sim, analisar para cada método, a diferença de classificação utilizando toda a amostra ou apenas as observações pertencentes a ZRG, os algoritmos LVQ e K-NN foram treinados com um número fixo de protótipos e de vizinhos, diferentemente do que foi feito na subseção 3.1.1.

Em relação à largura do sigma, utilizamos quatro abordagens diferentes: as primeiras duas baseadas na abordagem proposta de [41]; (i) validação cruzada; (ii) largura ótima supondo que as observações têm distribuição normal com variância unitária e o kernel é normal (equação 4.14 de [41]); (iii) valor médio de (i) e (ii) e (iv) Valor fixo (0.1). Para cada experimento, utilizamos o sigma que forneceu o melhor equilíbrio entre desempenho e número de observações selecionadas.

São apresentados 9 experimentos: **Experimento 10:** Classes divididas através da função cosseno, descrito na subseção 2.4.1. As ZRG's para este experimento podem ser encontradas nas Figuras 13-16; **Experimento 11:** Círculos concêntricos, descrito na subseção 2.4.3. As ZRG's para este experimento podem ser encontradas nas Figuras 17-20; **Experimento 12:** Waveform, descrito na subseção 2.4.4; **Experimento 13:** Letter-B, descrito na subseção 2.4.5; **Experimento 14:** Statlog, descrito na subseção 2.4.6;

Experimento 15: Diagnóstico de doenças cardíacas, descrito na subseção 2.4.7; **Experimento 16:** Classificação de punção aspirativa por agulha fina de nódulo mamário para diagnóstico, descrito na subseção 2.4.8; **Experimento 17:** Ionosfera, descrito na subseção 2.4.9; **Experimento 18:** Câncer de Pulmão, descrito na subseção 2.4.10.

Para os experimentos 10 a 14, utilizamos para LVQ e RN a média de 10 rodadas independentes. Para SVM, a largura do kernel e da constante C foram ambas variadas entre 2^0 , $2^{0.5}$, 2, 2^2 , 2^3 , 2^4 e 2^5 . Foram tomados os valores fora-da-amostra correspondentes ao melhor desempenho dentro-da-amostra. Em relação à k vizinhos mais próximos, utilizou-se $k = 3$.

Para os experimentos 15 a 17 (com menos observações), nós usamos validação cruzada (k-fold) com $k = 10$. Para cada banco de dados, uma partição foi criada e, para cada técnica (LVQ, RN, SVM e K-NN) a fase dentro-da-amostra foi efetuada por um conjunto de treinamento formado por 9 de 10 folds da partição. O fold remanescente foi usado para calcular o desempenho fora-da-amostra. Este procedimento foi repetido 10 vezes, em cada um deles com um fold diferente como conjunto de teste. Para SVM, em cada rodada da validação cruzada, houve variação da combinação dos parâmetros. Foi levado em consideração, para avaliação do desempenho fora-da-amostra, a configuração de parâmetros cuja validação cruzada apresentou o melhor desempenho dentro-da-amostra. Para o experimento 18 foi realizado um procedimento de leave-one-out com as 27 observações.

Na Tabela 3 encontra-se o tempo necessário para a obtenção das ZRG's. Na segunda coluna (denotada por Limpeza) encontra-se o tempo do procedimento de limpeza de cada banco de dados. Na coluna 3, o tempo para a obtenção dos sigmas 1 e 3. Estes valores coincidem porque o sigma 3 é a média entre os sigmas 1 e 2. Os tempos para os sigmas 2 e 4 não foram apresentados por se tratarem apenas de uma atribuição numérica, sendo desconsiderado o tempo de processamento.

Os resultados podem ser encontrados nas Tabelas 4, 6, 7 e 8. Também mostramos uma comparação com os métodos IB3 e DROP3 (Tabela 9). Estes dois algoritmos possuem um bom trade-off entre desempenho e quantidade de observações selecionadas. Na Tabela 5 encontra-se, ainda, uma comparação com a zona de risco proposta em [2]. Esta tabela deve ser comparada exclusivamente

com a Tabela 4, porque o método proposto em [2] é aplicável apenas para LVQ. Outra forma de medir o desempenho da RN foi apresentada na Tabela 10, através da média de valores da área sob a curva ROC, também conhecida como AUC [54]. Entre os 8 experimentos (o experimento 18 não foi considerado em função do procedimento de leave-one-out), a AUC da RN com ZRG perde em apenas 4 experimentos para AUC da RN treinada com todas as observações.

Tabela 3: Custo computacional em segundos para procedimento de limpeza, obtenção dos sigmas e ZRG's.

	Limpeza	Sigmas 1 e 3	ZRG 1	ZRG 2	ZRG 3	ZRG 4
Exp 10	3,4	120,6	11,3	9,3	7,4	7,2
Exp 11	5,2	300,4	16,9	15,7	14,9	14,7
Exp 12	33,0	2733,5	165,8	166,7	161,4	163,9
Exp 13	177,7	27316,9	1022,1	1033,0	1025,8	1014,6
Exp 14	66,6	7614,3	434,0	428,5	432,8	430,3
Exp 15	0,7	24,1	1,3	1,5	1,6	1,6
Exp 16	0,6	22,7	2,2	1,9	2,0	2,0
Exp 17	0,3	21,4	1,1	1,1	1,1	1,2
Exp 18	0,02	1,8	0,1	0,1	0,1	0,1

Tabela 4: Taxa de acerto fora-da-amostra para LVQ (parênteses indicam dentro-da-amostra).

	Todas as Observações				Observações da ZRG			
	Acerto	dp	(%) de obs.	Tempo	Acerto	dp	(%) de obs.	Tempo
Exp 10	97,9% (98,6%)	0,5 0,3	100%	20,3	98,1% (97,9%)	0,3 0,5	40,1%	8,0
Exp 11	97,8% (98,0%)	0,2 0,2	100%	27,1	99,2% (97,8%)	0,2 0,9	9,1%	2,4
Exp 12	91,6% (92,8%)	0,3 0,4	100%	34,6	91,4% (96,9%)	0,2 0,3	77,2%	25,9
Exp 13	96,2% (97,5%)	0,2 0,1	100%	127,0	94,6% (95,4%)	0,7 0,2	27,9%	31,0
Exp 14	98,0% (98,2%)	0,3 0,2	100%	51,5	98,4% (98,5%)	0,2 0,2	44,2%	21,4
Exp 15	78,9% (83,3%)	4,7 0,6	100%	6,5	79,5% (95,0%)	5,0 1,3	50,6%	3,2
Exp 16	97,1% (97,4%)	2,2 0,4	100%	5,9	96,9% (99,3%)	2,1 0,2	59,6%	3,5
Exp 17	87,2% (90,5%)	5,7 1,1	100%	3,2	86,0% (98,2%)	4,6 0,7	73,0%	2,2
Exp 18	85,2% (96,7%)	-	100%	0,2	81,5% (96,0%)	-	84,6%	0,2

Tabela 5: Taxa de acerto fora-da-amostra para LVQ (Zona de Risco).

	Observações da ZR			
	Acerto	dp	(%) de obs.	Tempo
Exp 10	96,4% (96,8%)	0,7 0,4	51,6%	297,2
Exp 11	97,6% (97,7%)	0,6 0,6	40,6%	400,6
Exp 12	83,4% (83,6%)	1,7 1,4	58,6%	578,5
Exp 13	77,7% (78,2%)	2,3 2,4	53,5%	3369,9
Exp 14	96,1% (95,8%)	1,4 1,3	55,0%	973,0
Exp 15	69,6% (73,3%)	7,2 3,3	58,4%	68,4
Exp 16	95,8% (96,2%)	2,4 0,9	49,4%	51,1
Exp 17	76,4% (79,0%)	11,2 7,2	40,1%	17,9
Exp 18	70,4% (74,4%)	-	55,4%	2,3

Tabela 6: Taxa de acerto fora-da-amostra para RN (parênteses indicam dentro-da-amostra).

	Todas as Observações				Observações da ZRG			
	Acerto	dp	(%) de obs.	Tempo	Acerto	dp	(%) de obs.	Tempo
Exp 10	99,4% (99,7%)	0,3 0,5	100%	6,1	99,5% (99,8%)	0,3 0,7	41,6%	4,0
Exp 11	99,7% (100%)	0,1 0,0	100%	2,7	99,5% (100%)	0,1 0,0	23,0%	0,8
Exp 12	89,7% (99,6%)	0,5 0,1	100%	76,1	90,2% (99,9%)	0,4 0,1	77,2%	62,8
Exp 13	98,6% (99,8%)	0,2 0,2	100%	77,7	97,7% (99,7%)	0,4 0,2	27,9%	23,9
Exp 14	98,8% (99,9%)	0,3 0,1	100%	87,5	96,2% (99,9%)	0,7 0,1	54,8%	62,4
Exp 15	74,9% (96,9%)	4,2 0,4	100%	6,1	77,2% (96,9%)	4,5 9,5	23,3%	2,0
Exp 16	94,9% (99,8%)	3,5 0,3	100%	5,2	95,2% (100%)	3,1 0,0	59,6%	3,7
Exp 17	88,0% (100%)	4,2 0,0	100%	14,5	86,0% (100%)	6,1 0,0	73,0%	3,8
Exp 18	85,2% (100%)	-	100%	44,5	74,1% (58,6%)	-	34,9%	10,9

Tabela 7: Taxa de acerto fora-da-amostra para SVM (parênteses indicam dentro-da-amostra).

	Todas as Observações				Observações da ZRG			
	Acerto	dp	(%) de obs.	Tempo	Acerto	dp	(%) de obs.	Tempo
Exp 10	99,0% (99,2%)	-	100%	1,0	85,8% (99,1%)	-	43,4%	0,3
Exp 11	99,0% (98,7%)	-	100%	2,4	99,5% (100%)	-	9,1%	0,1
Exp 12	67,6% (100%)	-	100%	5,8	67,7% (100%)	-	0,1%	0,01
Exp 13	99,4% (100%)	-	100%	23,6	98,8% (100%)	-	27,9%	2,5
Exp 14	98,3% (100%)	-	100%	4,7	98,1% (100%)	-	44,2%	1,1
Exp 15	75,0% (100%)	5,1 0,0	100%	0,3	79,2% (100%)	3,4 0,0	50,6%	0,1
Exp 16	95,9% (100%)	3,1 0,0	100%	5,5	81,5% (100%)	14,4 0,0	52,8%	2,2
Exp 17	89,5% (100%)	8,0 0,0	100%	4,3	66,7% (100%)	6,9 0,0	73,0%	2,2
Exp 18	70,4% (100%)	-	100%	0,01	74,1% (100%)	-	34,9%	0,01

Tabela 8: Taxa de acerto fora-da-amostra para K-NN (parênteses indicam dentro-da-amostra).

	Todas as Observações				Observações da ZRG			
	Acerto	dp	(%) de obs.	Tempo	Acerto	dp	(%) de obs.	Tempo
Exp 10	99,0% (99,0%)	-	100%	5,7	99,0% (99,7%)	-	16,2%	1,2
Exp 11	99,5% (99,3%)	-	100%	9,1	99,4% (98,4%)	-	9,1%	1,3
Exp 12	88,0% (86,0%)	-	100%	48,4	89,8% (97,2%)	-	75,6%	35,7
Exp 13	99,4% (99,4%)	-	100%	301,5	99,1% (99,3%)	-	27,9%	28,8
Exp 14	99,6% (98,9%)	-	100%	88,8	99,2% (99,5%)	-	44,2%	27,7
Exp 15	76,5% (76,4%)	3,4 0,4	100%	0,7	78,9% (96,4%)	3,5 0,6	50,6%	0,4
Exp 16	96,6% (96,9%)	2,5 0,4	100%	0,9	96,6% (98,1%)	2,2 1,6	52,8%	0,3
Exp 17	83,8% (84,7%)	5,7 1,1	100%	0,3	82,1% (97,5%)	6,0 0,8	34,1%	0,1
Exp 18	85,2% (84,6%)	-	100%	0,02	85,2% (94,9%)	-	84,6%	0,01

Tabela 9: Experimentos com IB3 e DROP3.

	IB3				DROP3			
	Acerto	dp	(%) de obs.	Tempo	Acerto	dp	(%) de obs.	Tempo
Exp 10	96,3%	0,8	3,9%	73,8	98,4%	-	5,2%	1297,1
Exp 11	98,8%	0,2	2,7%	118,7	99,4%	-	1,8%	2826,7
Exp 12	86,1%	0,9	8,2%	344,9	88,7%	-	9,7%	2312,6
Exp 13	94,9%	0,4	3,0%	2502,7	98,5%	-	4,7%	91382,2
Exp 14	98,4%	0,4	2,1%	189,5	98,8%	-	2,9%	9571,8
Exp 15	74,6%	6,4	14,5%	26,2	78,2%	3,5	11,1%	53,2
Exp 16	95,6%	2,4	2,2%	5,4	96,6%	2,1	2,9%	94,3
Exp 17	85,8%	4,6	12,9%	6,0	84,6%	5,3	6,0%	17,7
Exp 18	55,6%	-	11,0%	0,1	70,4%	-	27,9%	0,5

Tabela 10: AUC para RN.

	Todas as Observações	Observações da ZRG
Exp 10	0,9996	0,9998
Exp 11	1	1
Exp 12	0,9621	0,9657
Exp 13	0,9818	0,9702
Exp 14	0,9995	0,9940
Exp 15	0,8146	0,8478
Exp 16	0,9793	0,9758
Exp 17	0,9025	0,8605

Pode-se considerar que a técnica ZRG mostrou bons resultados quando aplicada a técnicas largamente utilizadas na literatura. O desempenho é em geral bom considerando uma razoável redução no número de observações utilizadas na fase dentro-da-amostra. Em relação a LVQ, a técnica mostrou resultados encorajadores nos experimentos 10, 11, 14 e 15, que podem ser conferidos na Tabela 4. Por exemplo, no experimento 11, um desempenho superior foi alcançado utilizando-se apenas 9,1% das observações. Já em comparação com a ZR proposta em [2] (Tabela 5), a utilização da ZRG para LVQ foi melhor em todos os 9 experimentos. Para RN e K-NN (Tabelas 6 e 8), a ZRG proveu uma redução considerável de observações com desempenhos comparáveis (melhores em alguns experimentos) em relação a todas as observações. Neste sentido, podemos considerar que a técnica foi capaz de encontrar subconjuntos apropriados da amostra nos quais o desempenho não caiu. Os destaques para RN foram os experimentos 15 e 16 nos quais resultados melhores foram atingidos utilizando-se respectivamente 23,3% e 59,6% das observações (Tabela 6). Já para K-NN (Tabela 8), podemos citar os experimentos 10, 15 e 16, onde resultados equivalentes ou melhores foram obtidos com 16,2%, 50,6% e 52,8% das observações. Para SVM, os melhores desempenhos relativos foram obtidos nos experimentos 15 e 18 (veja a Tabela 7), ganhos em torno de 4% na taxa de acerto com pouco mais da metade das observações (experimento 15) e 34,9% das observações (experimento 18).

Em relação a IB3 e DROP3, tratam-se de abordagens completamente diferentes do ponto de vista teórico. Experimentalmente, IB3 e DROP3 mostraram em geral uma redução significativa no número de observações, ainda mantendo

bom desempenho. A ZRG também mostrou reduções importantes no número de observações como pode ser observado nas Tabelas 4, 6, 7 e 8, mas não tanto quanto IB3 e DROP3 (Tabela 9). Entretanto, em relação à taxa de acerto, a ZRG foi melhor que IB3 em 26 dos 36 experimentos (4 métodos vezes 9 experimentos) e que o DROP3 em 20 dos 36.

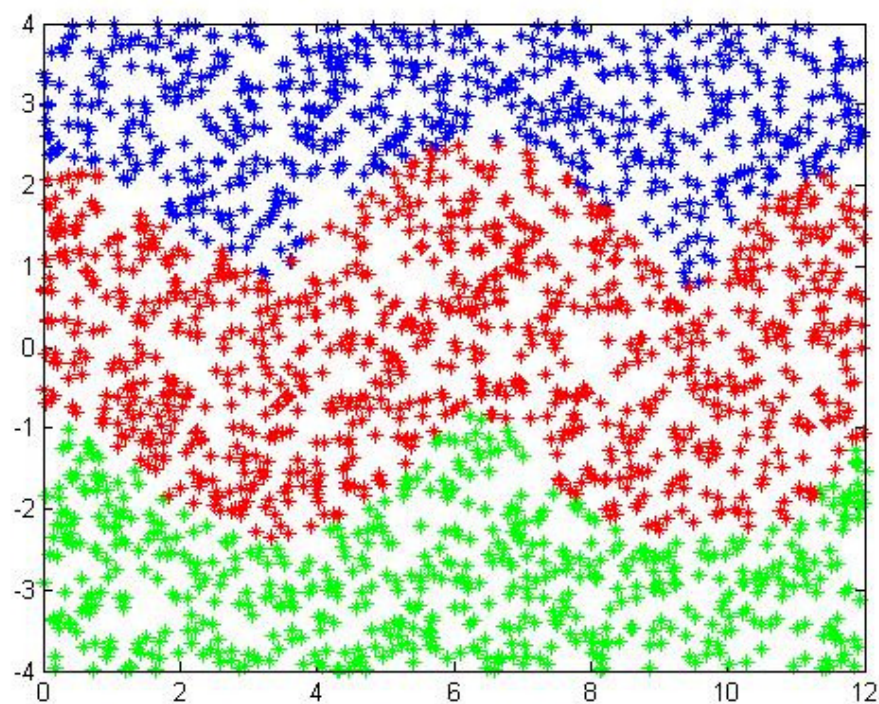


Figura 13: ZRG para o banco de dados apresentado na seção 2.4.1. Sigma obtido pela validação cruzada.

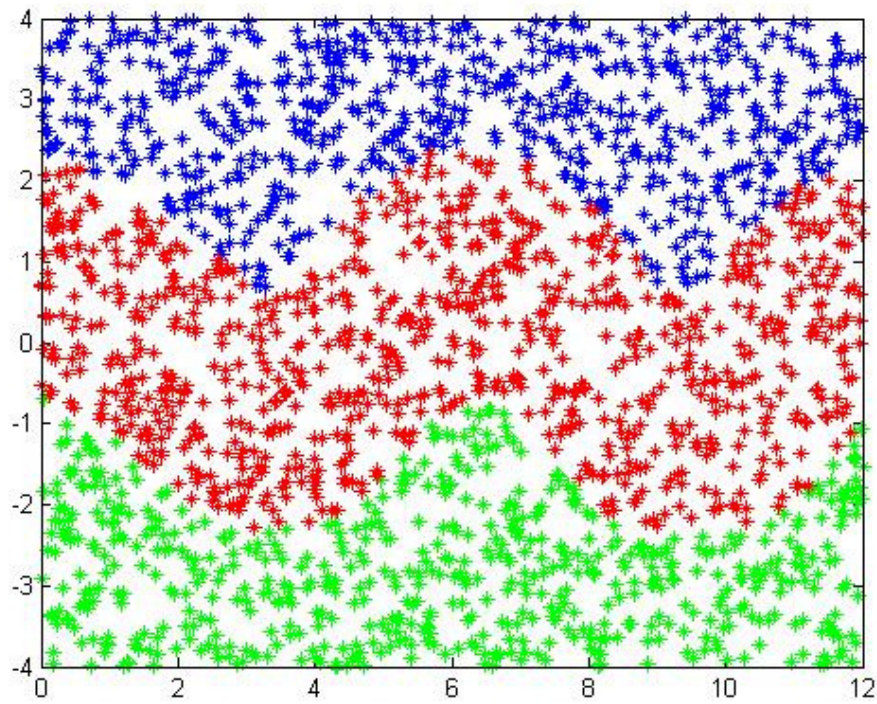


Figura 14: ZRG para o banco de dados apresentado na seção 2.4.1. Sigma obtido pela equação 4.14 de [41].

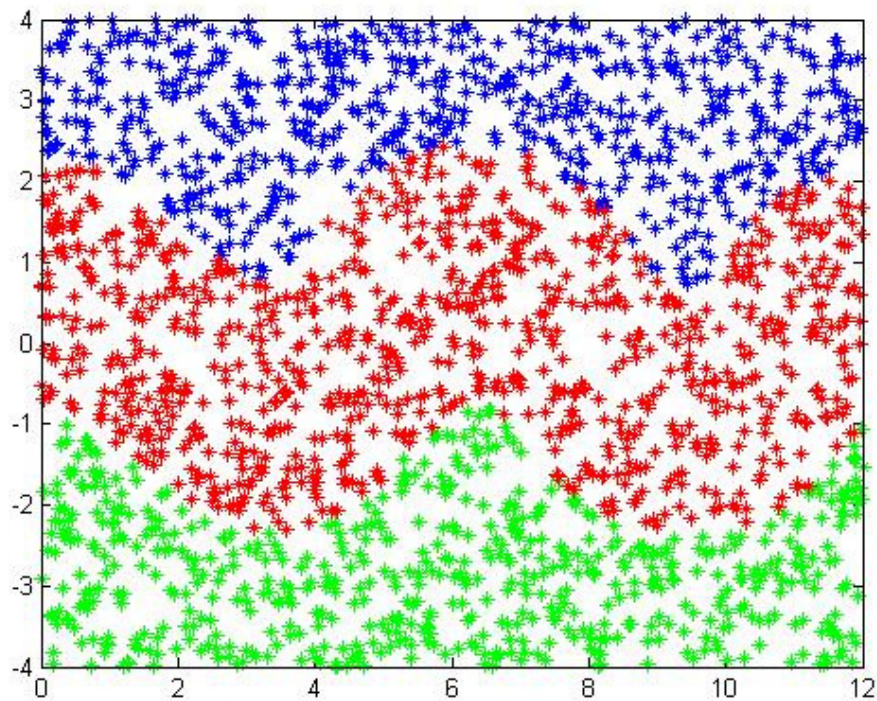


Figura 15: ZRG para o banco de dados apresentado na seção 2.4.1. Sigma obtido pela média entre os sigmas de validação cruzada e equação 4.14 de [41].

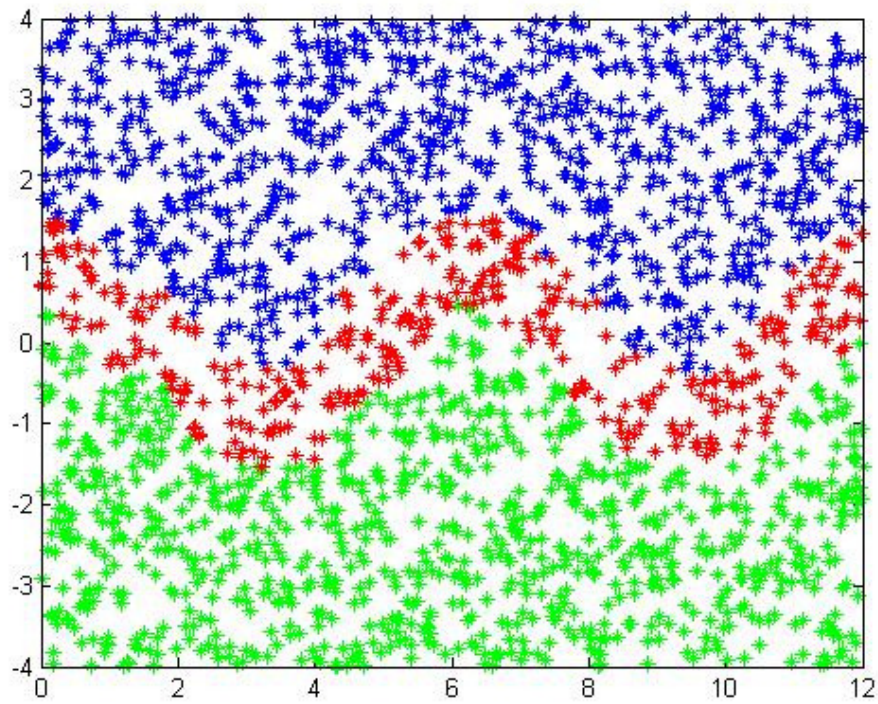


Figura 16: ZRG para o banco de dados apresentado na seção 2.4.1. Sigma 0.1.

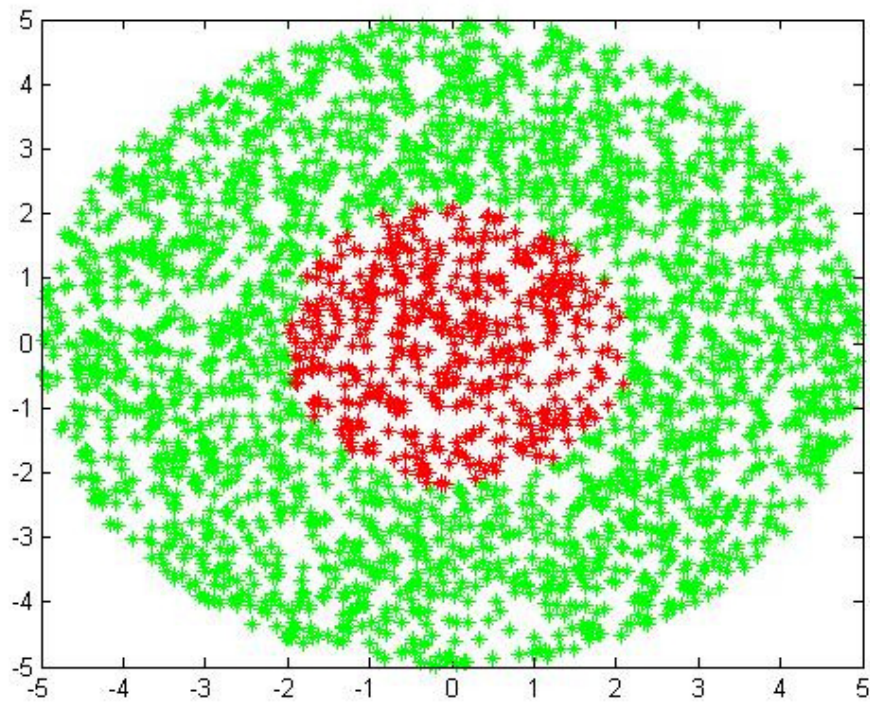


Figura 17: ZRG para o banco de dados apresentado na seção 2.4.3. Sigma obtido pela validação cruzada.

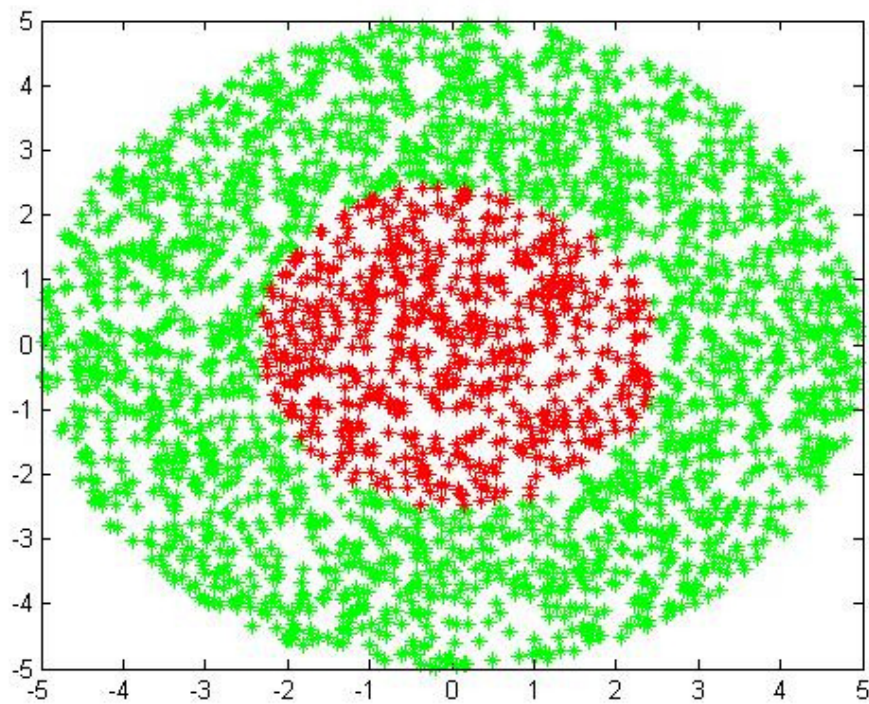


Figura 18: ZRG para o banco de dados apresentado na seção 2.4.3. Sigma obtido pela equação 4.14 de [41].

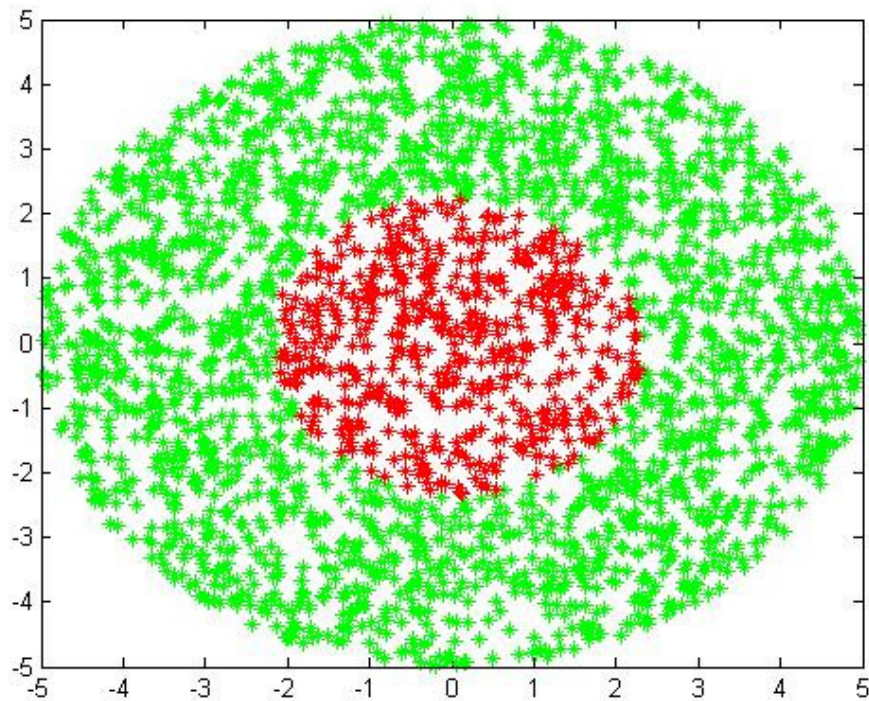


Figura 19: ZRG para o banco de dados apresentado na seção 2.4.3. Sigma obtido pela média entre os sigmas de validação cruzada e equação 4.14 de [41].

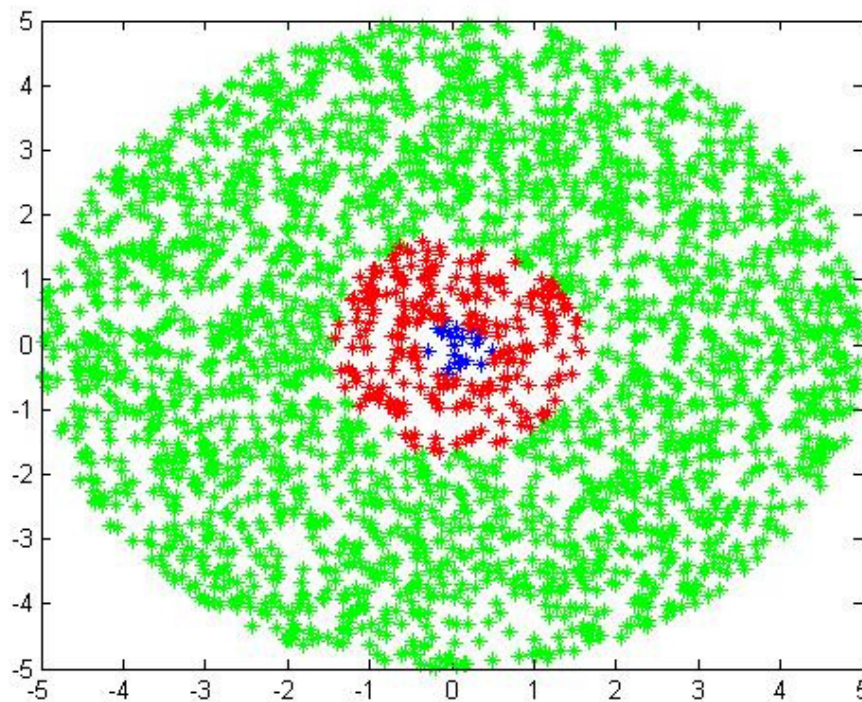


Figura 20: ZRG para o banco de dados apresentado na seção 2.4.3. Sigma 0.1.

3.2.

Resultados referentes à Seleção de Observações

Nesta seção estão apresentados os experimentos referentes às seções 2.1 e 2.2.1, onde foram apresentados métodos para seleção de observações.

3.2.1.

Uma Aplicação do Classificador por Quantização Vetorial (CQV) à Seleção de Observações

Nesta subseção apresentam-se os resultados correspondentes ao método CQV aplicado à seleção de observações, descrito no final da seção 2.1.

Os seguintes parâmetros, referentes ao algoritmo LBG, foram utilizados para todos os experimentos: $\zeta = 10^{-1}$ e $\varepsilon = 10^{-2}$.

Experimento 19: Classes divididas através da função cosseno, descrito na subseção 2.4.1. Neste experimento, vinte observações, em cada uma das classes, (cerca de 2% do total de observações) tiveram seu rótulo deliberadamente invertido. As observações cujos rótulos foram invertidos foram sorteadas de

acordo com uma distribuição uniforme. Na Figura 21, representam-se as observações das classes C_1 e C_2 por asteriscos, as com rótulos invertidos são representadas através de círculos. Os resultados encontram-se na Tabela 11. Por observações erroneamente selecionadas entendem-se as observações que foram selecionadas pelo método, mas que não tiveram seu rótulo invertido. Foram utilizados 128 protótipos. Na Tabela 12 encontram-se os resultados para DROP3 e IB3.

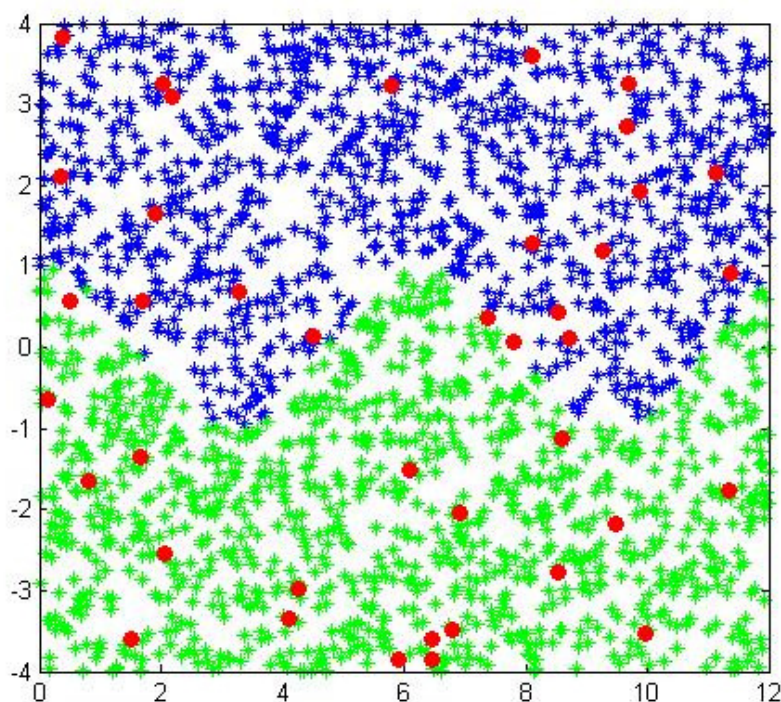


Figura 21: Classes C_1 e C_2 divididas através de função cosseno. Para cada classe, 20 observações (círculos vermelhos) tiveram seu rótulo invertido.

Tabela 11: Experimento 19: Reconhecimento de rótulos invertidos.

Classes	Rótulos Invertidos Identificados	Observações Erroneamente Selecionadas
C₁	19 entre 20 (95%)	6
C₂	20 entre 20 (100%)	7
Total	39 entre 40 (97.5%)	13 entre 2017 (0.6%)

Tabela 12: Experimento 19: Reconhecimento de rótulos invertidos por DROP3 e IB3.

Classes	Rótulos Invertidos Identificados DROP3	Rótulos Invertidos Identificados IB3
C₁	20 entre 20 (100%)	20 entre 20 (100%)
C₂	18 entre 20 (90%)	20 entre 20 (100%)
Total	38 entre 40 (95%)	40 entre 40 (100%)

Experimento 20: Círculos concêntricos, descrito na subseção 2.4.3. Foram utilizados 256 protótipos. Neste experimento, os rótulos de cinco observações foram invertidos na classe C₁ e de vinte observações na classe C₂ (Figura 22). Os resultados estão na Tabela 13. Na Tabela 14 encontram-se os resultados para DROP3 e IB3.

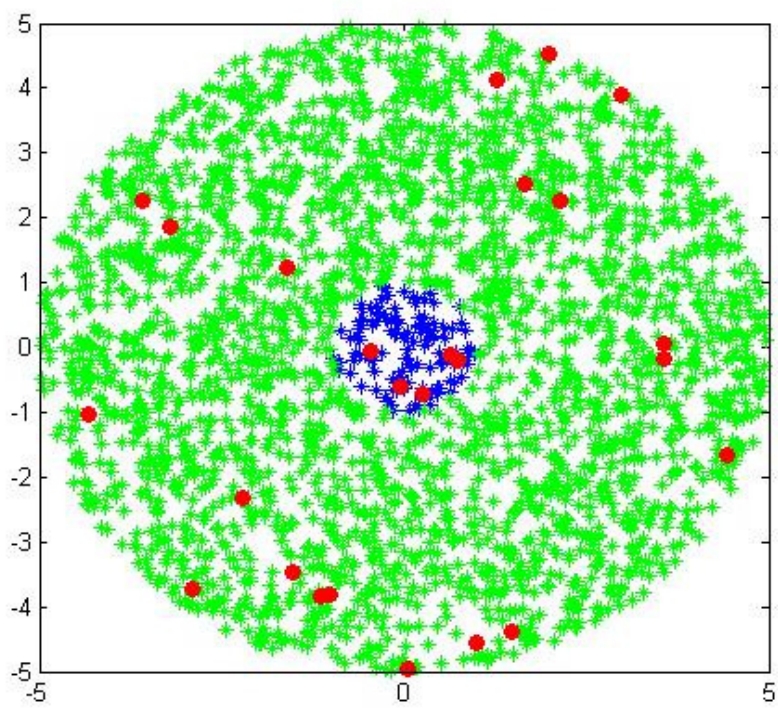


Figura 22: Classes C₁ e C₂ – Círculos Concêntricos com 5 e 20 observações com rótulo invertido (círculos vermelhos).

Tabela 13: Experimento 20: Reconhecimento de rótulos invertidos.

Classes	Rótulos Invertidos Identificados	Observações Erroneamente Seleccionadas
C ₁	5 entre 5 (100%)	2
C ₂	20 entre 20 (100%)	10
Total	25 entre 25 (100%)	12 entre 2709 (0.4%)

Tabela 14: Experimento 20: Reconhecimento de rótulos invertidos por DROP3 e IB3.

Classes	Rótulos Invertidos Identificados DROP3	Rótulos Invertidos Identificados IB3
C₁	5 entre 5 (100%)	4 entre 5 (80%)
C₂	20 entre 20 (100%)	18 entre 20 (90%)
Total	25 entre 25 (100%)	22 entre 25 (88%)

Experimento 21: Quadrados Sobrepostos, descrito na subseção 2.4.2. Foram utilizados 32 protótipos. Neste banco, vinte observações tiveram seu rótulo invertido em cada classe (2% das observações por classe), como pode-se observar na Figura 23. Os resultados estão na Tabela 15. O método proposto selecionou 249 observações (112 da classe C_1 e 137 da classe C_2) com a maioria pertencendo à interseção entre os dois quadrados, além de 36 observações que tiveram seus rótulos invertidos. As observações que não foram retiradas pelo método estão mostradas na Figura 24. Resultados para DROP3 e IB3 na Tabela 16.

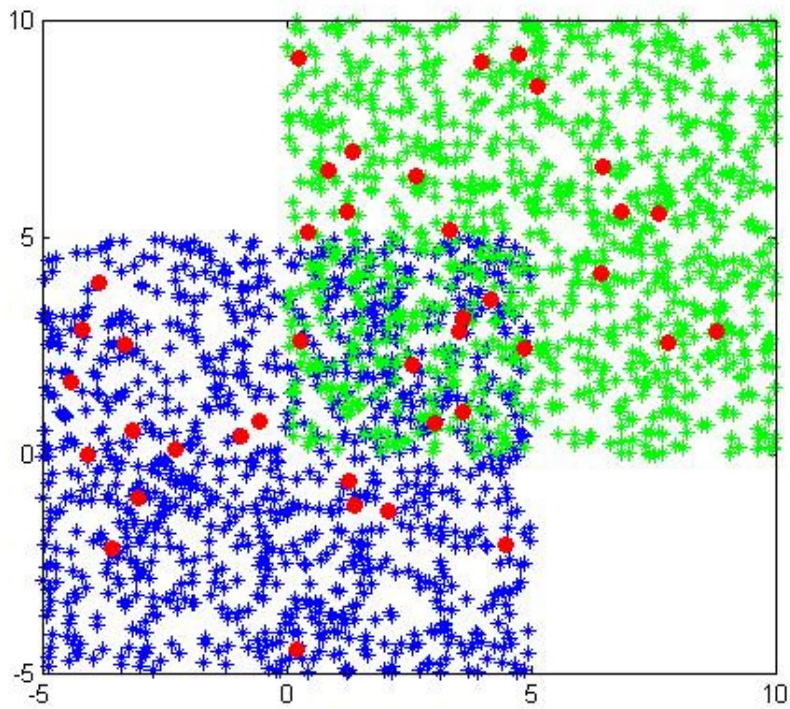


Figura 23: Classes C_1 e C_2 – Quadrados Sobrepostos. Para cada classe, 20 observações (círculos vermelhos) tiveram seu rótulo invertido.

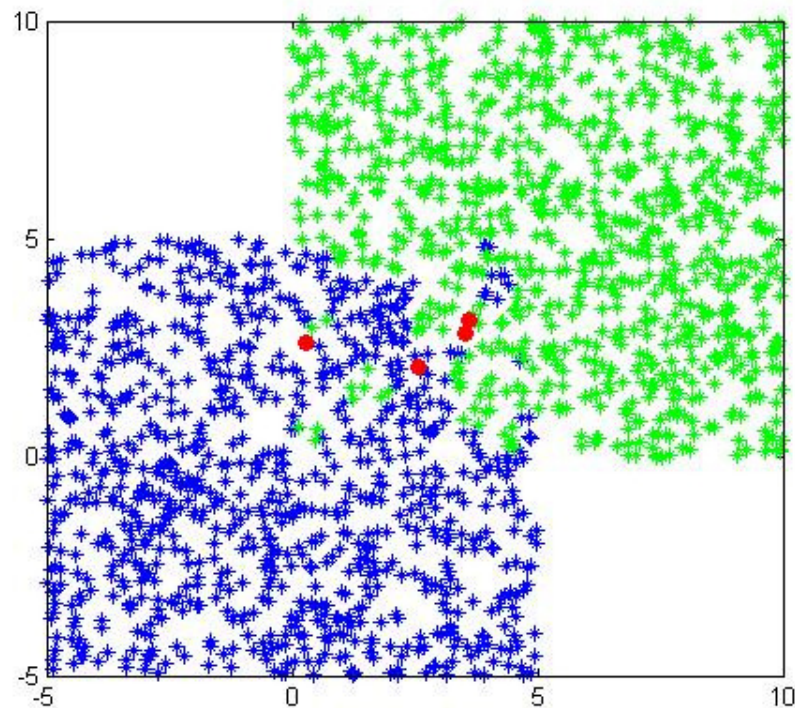


Figura 24: Classes C_1 e C_2 – Quadrados Sobrepostos. Observações mantidas pelo método CQV Aplicado a Seleção de Observações.

Tabela 15: Experimento 21: Reconhecimento de rótulos invertidos.

	Rótulos Invertidos Identificados
C₁	18 entre 20 (90%)
C₂	18 entre 20 (90%)
Total	36 entre 40 (90%)

Tabela 16: Experimento 21: Reconhecimento de rótulos invertidos por DROP3 e IB3.

Classes	Rótulos Invertidos Identificados DROP3	Rótulos Invertidos Identificados IB3
C₁	20 entre 20 (100%)	20 entre 20 (100%)
C₂	19 entre 20 (95%)	20 entre 20 (100%)
Total	39 entre 40 (97.5%)	40 entre 40 (100%)

Note-se que, o CQV aplicado à seleção de observações é eficiente não só para selecionar observações com rótulos invertidos, mas também para identificar regiões em que a geração das observações pode ser aproximadamente uniforme (Figura 24). Esta característica pode ser particularmente interessante para aplicações em problemas de classificação, onde observações pertencentes a uma região uniforme poderiam ser retiradas antes da fase de treinamento. Em relação à identificação de rótulos invertidos, o procedimento obteve sucesso, já que nos 3 experimentos foram identificados respectivamente 97,5%, 100% e 90% das observações com rótulos invertidos (Tabelas 11, 13 e 15). Nos bancos de dados que não possuíam regiões uniformes a taxa de observações erroneamente selecionadas foi irrelevante (0,6% no banco 2.4.1 e 0,4% no banco 2.4.3; Tabelas

11 e 13). Já a comparação com os métodos DROP3 e IB3 (Tabelas 12, 14 e 16) mostram a competitividade do método proposto em comparação a estes dois métodos clássicos.

3.2.2.

Resultados da Aplicação de Zona de Risco Generalizada (ZRG) em Seleção de Observações

Experimentos referentes à metodologia apresentada na seção 2.2.1. Trata-se de um método para classificar observações como típicas, de risco ou com rótulo invertido, baseado no contexto de ZRG. Os valores de sigma utilizados são: VC: correspondente a validação cruzada; Plug-in: correspondente a equação 4.14 de [41], Média: correspondente a média de VC e Plug-in e Valor fixo 0.1.

Experimento 22 (Classes divididas através da função cosseno; descrito na seção 2.4.1)

Vinte observações tiveram seu rótulo invertido em cada classe. Na Tabela 17 encontram-se as classificações das amostras em típicas (T), de risco (R) e com inversão de rótulo (Inv) para diferentes larguras da função de *kernel*. Na Tabela 18, pode-se encontrar a quantidade de observações que tiveram o rótulo invertido reconhecidas pelo método (classificadas como Inv) e a quantidade de observações classificadas como Inv erradamente, isto é, sem terem tido o rótulo invertido.

Tabela 17: Observações classificadas em típicas (T), de risco (R) ou com inversão de rótulo (Inv). Entre parênteses, a porcentagem que representam em relação ao total de observações.

	VC	Plug-in	Média	0.1
T	241	295	267	1339
R	1754	1704	1729	681
Inv	62	58	61	37

Tabela 18: Observações classificadas como Inv que de fato tiveram o rótulo invertido e observações classificadas como Inv que não tiveram o rótulo invertido.

	Rótulos Invertidos Selecionados como Inv (no máximo 40)	Rótulos Não Invertidos Selecionados como Inv	Total
VC	38	24	62
Plug-in	38	20	58
Média	38	23	61
0.1	35	2	37

Experimento 23 (Círculos Concêntricos – Observações dentro-da-amostra; descrito na seção 2.4.3)

Cinco observações tiveram seu rótulo invertido na classe 1 e vinte na classe 2. Na Tabela 19 encontram-se as classificações das observações e, na Tabela 20, as observações que tiveram o rótulo invertido reconhecidas pelo método e a quantidade de observações classificadas como Inv erradamente.

Tabela 19: Observações classificadas em típicas (T), de risco (R) ou com inversão de rótulo (Inv). Entre parênteses, a porcentagem que representam em relação ao total de observações.

	VC	Plug-in	Média	0.1
T	1249	1255	1223	2131
R	1379	1341	1391	561
Inv	106	138	120	42

Tabela 20: Observações classificadas como Inv que de fato tiveram o rótulo invertido e observações classificadas como Inv que não tiveram o rótulo invertido.

	Rótulos Invertidos Selecionados como Inv (no máximo 25)	Rótulos Não Invertidos Selecionados como Inv	Total
VC	25	81	106
Plug-in	25	113	138
Média	25	95	120
0.1	8	34	42

Experimento 24 (Quadrados Sobrepostos; descrito na seção 2.4.2)

Neste experimento não houve inversão de rótulo. Na Tabela 21 encontram-se as classificações das observações. Na Figura 25, o resultado da seleção de observações para sigma obtido pela validação cruzada.

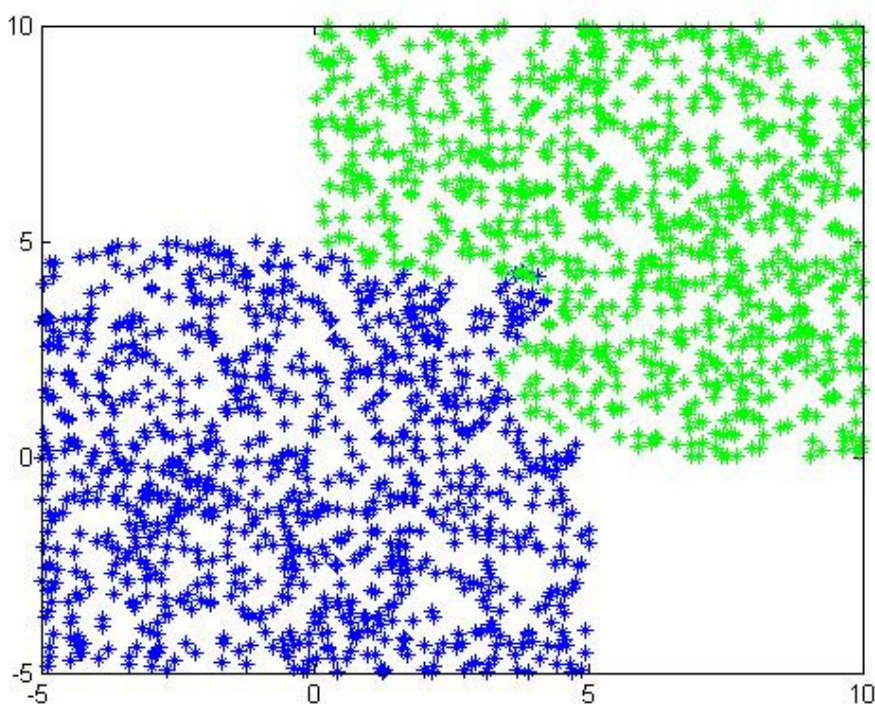


Figura 25: Classes C_1 e C_2 – Quadrados Sobrepostos. Observações mantidas pelo método ZRG Aplicado a Seleção de Observações.

Tabela 21: Observações classificadas em típicas (T), de risco (R) ou com inversão de rótulo (Inv). Entre parênteses, a porcentagem que representam em relação ao total de observações.

	VC	Plug-in	Média	0.1
T	1310	1419	1377	1486
R	436	368	374	424
Inv	254	213	249	90

Experimento 25 (Quadrados Sobrepostos; descrito na seção 2.4.2)

Vinte observações tiveram seu rótulo invertido em cada classe. Na Tabela 22 encontram-se as classificações das observações e, na Tabela 23, as observações que tiveram o rótulo invertido reconhecidas pelo método e a quantidade de observações classificadas como rótulo invertido (Inv) erradamente.

Tabela 22: Observações classificadas em típicas (T), de risco (R) ou com inversão de rótulo (Inv). Entre parênteses, a porcentagem que representam em relação ao total de observações.

	VC	Plug-in	Média	0.1
T	1288	1406	1358	1428
R	425	352	360	462
Inv	287	242	282	110

Tabela 23: Observações classificadas como Inv que de fato tiveram o rótulo invertido e observações classificadas como Inv que não tiveram o rótulo invertido.

	Rótulos Invertidos Selecionados como Inv (no máximo 40)	Rótulos Não Invertidos Selecionados como Inv	Total
VC	34	253	287
Plug-in	39	203	242
Média	36	246	282
0.1	22	88	110

Em termos de reconhecimento de rótulos invertidos reconhecidos, tem-se que a aplicação de ZRG a seleção de observações mostrou-se competitiva com os métodos CQV, DROP3 e IB3 quando utilizados com o mesmo objetivo.