

IA004

Relatório do Segundo Trabalho

Alunos:

Arthur Zanatta da Costa - RA: 116194

Sérgio Vieira Bueno - RA: 072372

Professor :

Romis Ribeiro de Faissol Attux

Nov

2016

Questão 1

a)

Para obter a máquina ótima de classificação, o primeiro passo foi traçar as verossimilhanças $P(x/C_i)$, $i = -1, 1$. A figura abaixo mostra as distribuições. Percebe-se que $P(x/C_{-1}) > P(x/C_{+1}) \leftrightarrow x \in [0, 1]$.

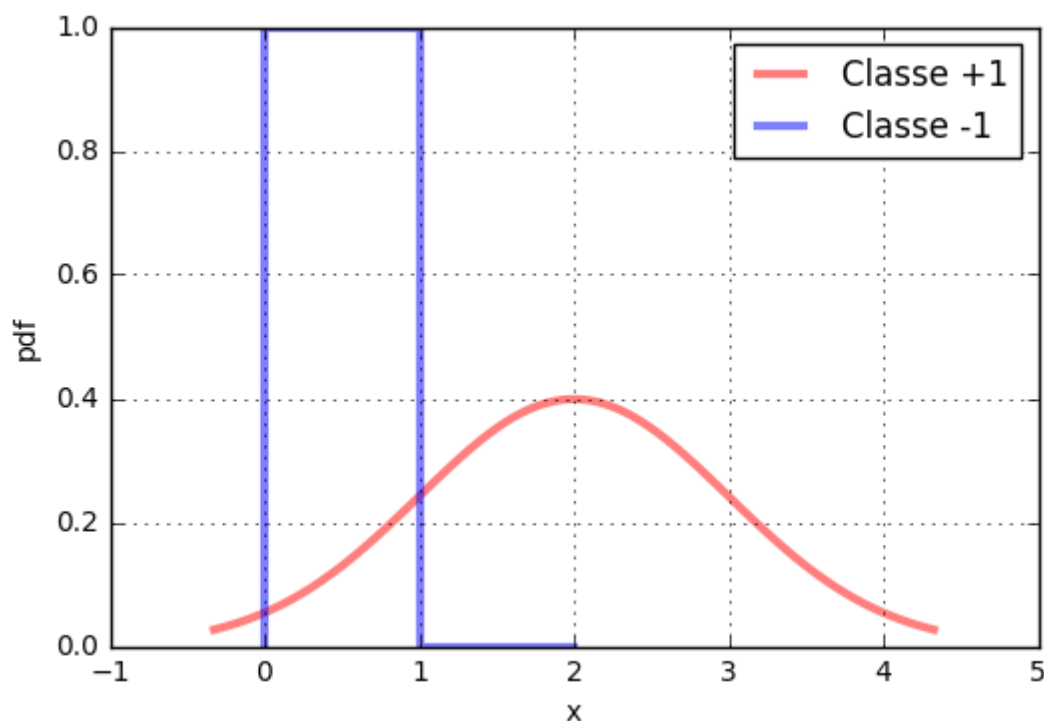


Figura 1: Verossimilhanças $P(x/C_i)$ das classes -1 e $+1$.

Em seguida, cada verossimilhança foi ponderada pela distribuição a priori da classe $P(C_i)$, $i = -1, 1$. A figura abaixo mostra as distribuições ponderadas. Percebe-se que $P(C_{-1})P(x/C_{-1}) > P(C_{+1})P(x/C_{+1}) \leftrightarrow x \in [0, 1]$. Ou seja, o intervalo de decisão se mantém o mesmo.

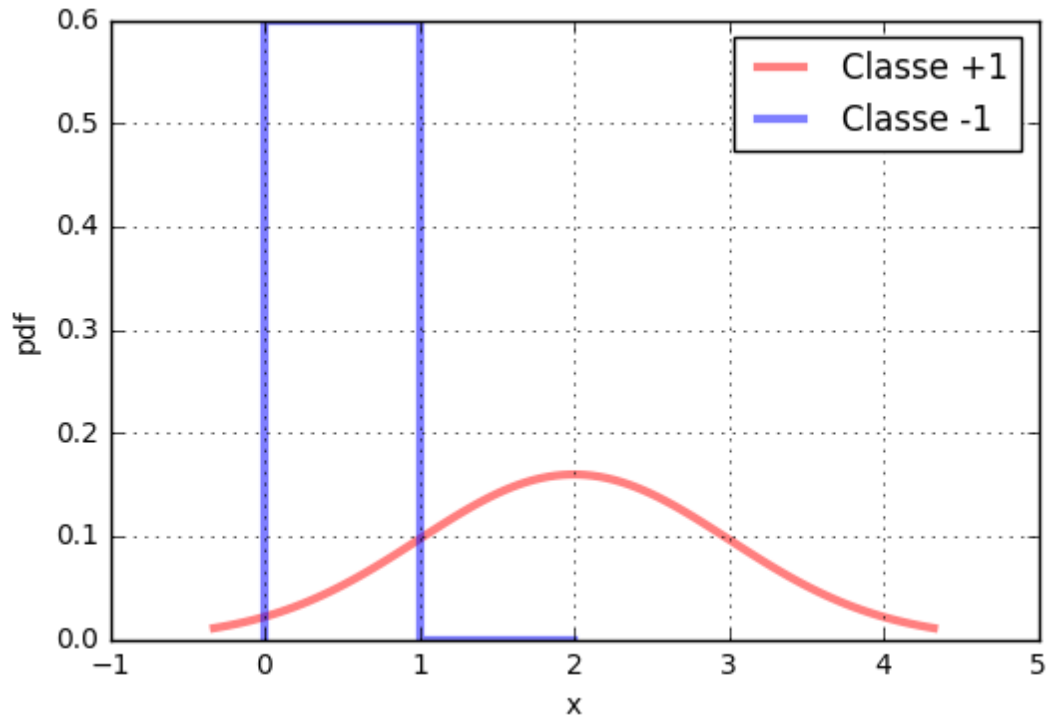


Figura 2: Verossimilhanças ponderadas $P(C_i)P(x/C_i)$ das classes -1 e $+1$.

Pelas figuras 1 e 2, o grupo concluiu que os intervalos de decisão da metodologia de Bayes MAP são $[0, 1]$ para classe C_{-1} e $\mathbb{R} \setminus [0, 1]$ para a classe C_{+1} .

O diagrama da máquina ótima de classificação está em anexo.

b)

A fronteira definida pelo MAP é composta pelos pontos do conjunto $\{0, 1\}$. Os intervalos de decisão são $[0, 1]$ para classe C_{-1} e $\mathbb{R} \setminus [0, 1]$ para a classe C_{+1} . Logo, A probabilidade de erro do classificador MAP é $P(\text{erro}) = 5,4\%$. A solução analítica desse problema está em anexo.

c)

A partir das figuras 1 e 2, é possível dizer que a fronteira de decisão é a mesma para o *MAP* e para o *ML*. Como a fronteira de decisão é a mesma, a integral do item b não se altera. Logo, a probabilidade de erro do classificador *ML* é $P(\text{erro}) = 5,4\%$.

d)

As duas metodologias analisadas apresentam o mesmo desempenho na tarefa de classificação deste problema. O método *ML* faz uso da falácia da taxa de base ao ignorar o prior durante os cálculos dos custos médios. Como consequência, ele costuma ser mais fácil de calcular, visto que não é preciso calcular ou estimar as probabilidades a priori. O preço a ser pago por essa comodidade costuma ser resultados inferiores aos do método *MAP*. No entanto, neste caso em particular o formato da verossimilhança da classe -1 fez com que os critérios tivessem a mesma performance. Isso porque a pdf uniforme apresenta valor constante superior a pdf gaussiana em seu intervalo de definição (figura 1) e essa relação se mantém após a ponderação pelos priors (figura 2).

Questão 2

a)

O grupo escolheu trabalhar com o problema de modelamento de qualidade de vinho baseado em testes psicoquímicos do repositório de Machine Learning da UC Irvine (<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>). Existem dois conjuntos de dados: vinho vermelho e vinho branco. Ambos são variações do “Vinho Verde” Português. Eles apresentam o mesmo número de atributos reais e as mesmas classes. Como o vinho branco apresenta um número maior de instâncias, o grupo resolveu trabalhar com ele.

As entradas são medidas de componentes e propriedades físico-químicas do vinho. A saída do problema é a qualidade do vinho. Ela pertence a um conjunto finito de inteiros e, por conseguinte, pode ser formulado como um problema de classificação.

É importante notar que a qualidade/nota do vinho, ao contrário dos atributos de entrada, é quantificada de forma pouco confiável e subjetiva. Cada nota é a mediana das notas conferidas por três degustadores distintos. Não é absurdo supor que existe uma relação entre ela e os atributos físico-químicos que constituem o vetor de entrada. Todavia, é bem improvável que esses atributos consigam fazer uma separação perfeita entre as notas. Afinal, nada impede de vinhos com propriedades físico-químicas próximas de idênticas recebam notas distintas. Essa informação foi levada em conta na hora de avaliar o desempenho dos classificadores e será mais detalhada ao longo do relatório.

Vale mencionar que este repositório é indicado pelos seus contribuintes para estudos de seleção de atributos, um dos temas desse exercício. Isso porque eles não sabem se todas os atributos iniciais são relevantes.

Explicado o problema, pode-se entrar em detalhes com respeito aos dados. O conjunto de dados é composto por 11 atributos com valores reais:

1. acidez fixa
2. acidez volátil
3. ácido cítrico
4. açúcar residual
5. cloretos
6. dióxido de enxofre livre
7. dióxido de enxofre total
8. densidade
9. pH
10. sulfatos
11. álcool

A saída é a qualidade do vinho. Ela é dada como uma nota entre 1 e 10, totalizando 10 classes para o problema. Entretanto, o conjunto não apresenta as 10 classes. Os dados referentes ao vinho branco contém notas de 3 a 9, totalizando 7 classes.

No processamento inicial dos dados, observou-se que em nenhuma das linhas de entrada falta informação a respeito dos atributos ou das notas. Também observou-se que as saídas não são balanceadas, ou seja, não apresentam a mesma frequência relativa. A figura 3 mostra o diagrama de Pareto do conjunto de dados. A tabela 1 mostra as informações das frequências relativas em mais detalhes. A tabela 2 sumariza algumas informações a respeito do conjunto de dados.



Figura 3: Diagrama de Pareto para o conjunto de dados do vinho branco.

Tabela 1: Frequência relativa de cada classe.

Qualidade	Quantidade	Frequência Relativa
3	20	0.41 %
4	163	3.3 %
5	1457	29.7 %
6	2198	44.9 %
7	880	18.0 %
8	175	3.6 %
9	5	1 %
Total	4898	100 %

Tabela 2: Resumo das informações do conjunto de dados.

Vinho branco	
Número de instâncias	4898
Número de atributos inicial	11
Número de classes	7
Saída balanceada	Não
Realizar seleção de atributos	Sim
Classes linearmente separáveis	Não se sabe

b)

O grupo escolheu realizar a divisão convencional entre dados de treinamento, validação e teste. Primeiro, separou-se 20% dos dados para teste. Os restantes foram divididos entre treino e validação na proporção 4 : 1, respectivamente. O grupo garantiu que as classes menos frequentes (3, 5, 7 e 8) estivessem presentes nessas exatas proporções nos três conjuntos. Houve a opção por não utilizar a validação cruzada, porque o grupo acredita que essa divisão gera dados suficientes para um bom treinamento e escolha de hiperparâmetros. A tabela abaixo mostra o número de instâncias para cada conjunto formado.

Tabela 3: Separação dos dados em treino, validação e teste.

Vinho branco	Número de instâncias
Conjunto de treino	3133
Conjunto de validação	785
Conjunto de teste	980

c)

Nesta etapa, o grupo usou uma metodologia de filtro para seleção de atributos usadas no problema. Escolheu-se o coeficiente de Pearson como técnica de filtragem. A escolha foi feita devido a simplicidade de implementação em comparação com outros métodos como a informação mútua. A dificuldade deste último método seria estimar as distribuições de probabilidade de forma acurada.

Calculando o coeficiente de Pearson de cada atributo e tomando seu módulo, chegou-se ao ranking de atributos mostrado na tabela 4.

Tabela 4: Ranking dos atributos em ordem decrescente.

Atributo	Módulo do coeficiente de Pearson
álcool	0.45
dióxido de enxofre total	0.17
cloretos	0.15
acidez fixa	0.12
pH	0.11
açúcar residual	0.10
acidez volátil	0.07
sulfatos	0.06
dióxido de enxofre livre	0.01
ácido cítrico	0.01
densidade	0.00

O grupo optou por trabalhar com os cinco atributos com os maiores módulos do coeficiente de Pearson. Portanto, a etapa de filtragem selecionou os seguintes atributos: álcool, dióxido de enxofre total, cloretos, acidez fixa, pH.

A figura abaixo mostra a matriz de espalhamento e o histograma dos atributos selecionados e da saída. As classes são separadas por cores distintas. Olhando os atributos dois a dois, percebe-se que não há nenhuma separação evidente entre as classes. Isso reforça mais uma vez a dificuldade desse problema de classificação onde a classe é uma nota calculada de forma subjetiva e pouco confiável.



Figura 4: Matriz de espalhamento e histogramas dos atributos selecionados e da saída. Não há nenhuma separação evidente entre as classes.

d)

Nessa etapa, o grupo projetou uma Extreme Learning Machine (ELM) com o intuito de classificar a qualidade do vinho de forma correta. A ELM foi projetada como um regressor com uma camada de saída, como visto em aula. O grupo assumiu que a classe predita é o inteiro mais próximo do valor de saída da rede.

Como hiperparâmetros da rede, o grupo colocou o número de neurônios da camada intermediária N , o intervalo da distribuição uniforme I e o coeficiente de regularização λ . O conjunto de validação foi usado para determinar os hiperparâmetros. O valor de N foi varrido de 8 até 512 neurônios, dobrando o número de neurônios em cada mudança. O valor de I foi varrido entre os valores (0.1, 1, 10). O valor de λ foi varrido entre 2^{-24} e 2^{+25} , dobrando o valor da regularização em cada mudança, como recomendado em aula.

Como métrica de desempenho, o grupo utilizou a acurácia. Ela mede a porcentagem de exemplos classificados corretamente. O treinamento foi realizado de maneira convencional. Os pesos da camada intermediária são amostras da distribuição uniforme. Os pesos da camada de saída são a solução do problema de mínimos quadrados, levando em conta o termo de regularização. A acurácia de treinamento foi calculada para verificar se houve a ocorrência de sob-ajuste (underfitting). A acurácia de validação foi usada para determinar qual combinação de parâmetros era a melhor. As figuras 5 e 6 mostram as acurácias de treinamento e validação, respectivamente. A partir da acurácia de validação, escolheu-se o conjunto de parâmetros 1030 como o melhor para o problema. Ele corresponde a $N = 512$, $I = 10$ e $\lambda = 64$ e provê uma acurácia de 53,5 % no conjunto de validação.

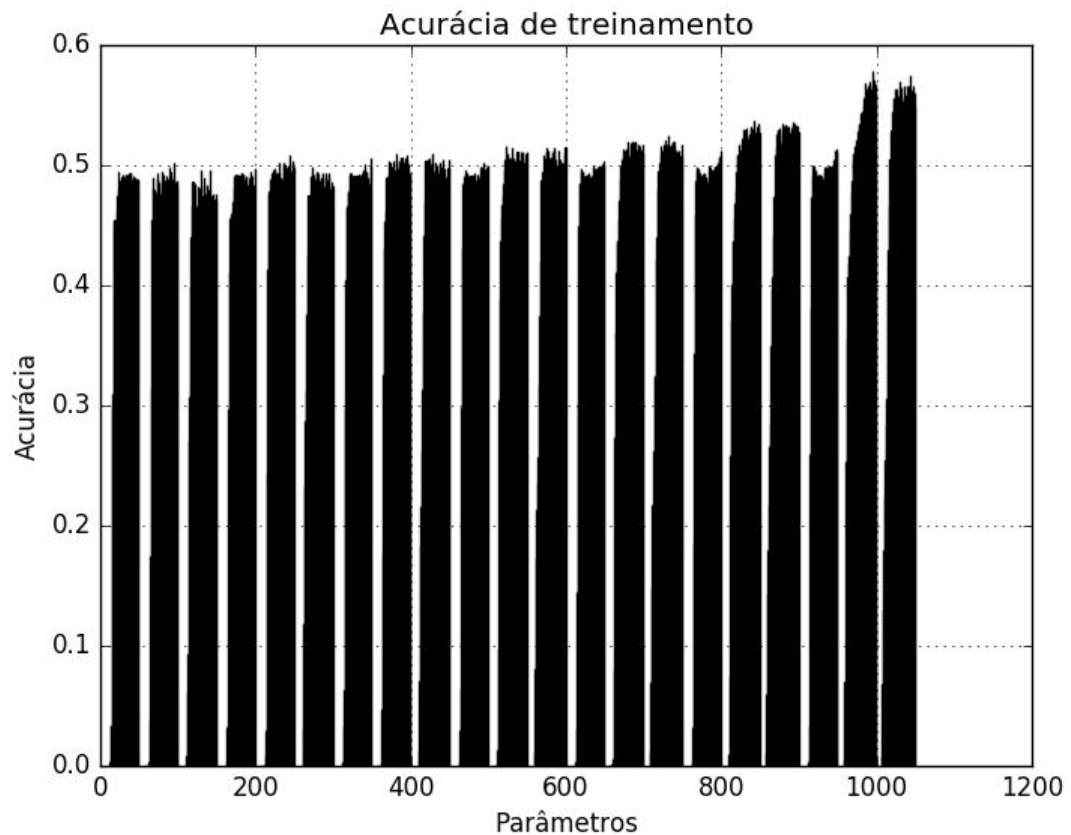


Figura 5: Acurácia do conjunto de treinamento no treino da ELM.

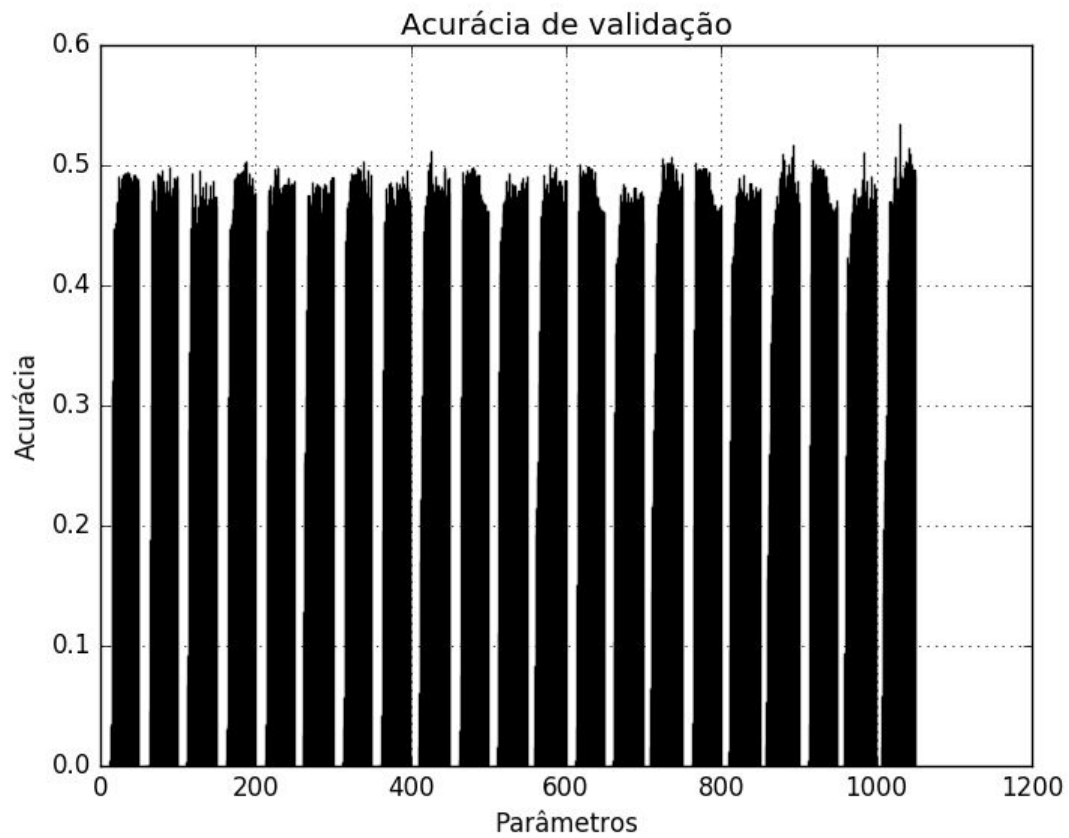


Figura 6: Acurácia do conjunto de validação no treino da ELM.

Com os melhores hiperparâmetros determinados, a rede foi treinada mais uma vez usando esses parâmetros e os conjuntos de treino e validação como entrada (essa combinação será chamada de conjunto de treino final, por conveniência). Em seguida, calculou-se a acurácia no conjunto de treino final e no conjunto de teste. Uma vez que os pesos entre os dados de entrada e a camada intermediária são aleatórios, rodou-se o algoritmo de treinamento algumas vezes. Escolhemos rodar 100 vezes e pegar o melhor resultado de acurácia de treino. Os resultados são mostrados na tabela abaixo.

Tabela 5: Acurácia dos conjuntos de treino final e de teste.

Acurácia	
Treino final	54,7 %
Teste	52,3 %

Percebe-se que o classificador tem uma performance bem medíocre na tarefa de classificar a qualidade do vinho. O grupo já suspeitava que isso fosse acontecer, devido a discussão sobre a forma de obtenção da qualidade do vinho feita no final do item '2a'. Certamente, classificar "corretamente" cada amostra de vinho é uma tarefa muito complicada e de certa forma até sem sentido, dado que a nota final é a mediana de três notas de degustadores distintos. Por exemplo, uma amostra com qualidade igual a 6, pode ter vindo de três degustadores que deram ao vinho as notas 5, 6 e 7, respectivamente. Ou pode ter vindo das notas 4, 6 e 7. Ou até mesmo de uma situação ideal onde as notas seriam 6, 6 e 6. Ao mesmo tempo, um vinho com propriedades físico-químicas (vetor de entrada do problema) semelhantes ao primeiro poderia receber as notas 5, 5 e 6, ficando com uma qualidade final de 5.

Como pessoas diferentes percebem o sabor de forma distinta, não existe um mapeamento bem definido entre os atributos físico-químicos e as notas dadas. Além disso, não está claro o significado de cada ponto de escala, o que, aumenta o espaço para interpretações diferentes entre os entrevistados¹. De certa forma, para o problema de classificação, teria sido melhor se os criadores do conjunto de dados tivessem pedido aos avaliadores para darem uma nota entre : 'Ruim', 'Médio' e 'Bom'². Afinal, é muito mais simples de se chegar a um consenso de que um vinho é ruim, do que chegar a um consenso de que a sua nota é 5, em uma escala de 1 a 10.

Apesar dessa dificuldade, é possível obter um resultado interessante ao mudar a métrica de desempenho do classificador. Já foi estabelecido que a qualidade do vinho não é uma medida absoluta. Mesmo assim, ela carrega uma informação importante consigo. A verdadeira qualidade do vinho deve ser um número próximo do valor de saída. Portanto, se o nosso classificador estima o valor 7 e a saída é 6, isso é muito melhor do que um classificador que estima o valor 1. Para observar nosso classificador sob essa ótica, foi calculada a matriz de confusão do classificador. Ela é mostrada na tabela 6. De fato, observa-se a maior parte dos erros a uma unidade de distância do valor considerado correto.

¹ Krosnick, Jon A., and Leandre R. Fabrigar. "Designing rating scales for effective measurement in surveys." *Survey measurement and process quality*(1997): Cap 6.

² Krosnick, Jon A., and Matthew K. Berent. "Comparisons of party identification and policy preferences: The impact of survey question format." *American Journal of Political Science* (1993): 941-964.

Tabela 6: Matriz de confusão da ELM.

	3	4	5	6	7	8	9
3	0	1	1	2	0	0	0
4	0	0	18	13	2	0	0
5	0	0	134	139	9	0	0
6	0	1	85	291	72	0	0
7	0	0	7	106	63	0	0
8	0	0	1	13	21	0	0
9	0	0	0	0	1	0	0

Graças a essa observação, o grupo definiu uma acurácia folgada. Ela considera que toda amostra em que a distância entre a saída e a estimacão é menor ou igual a 1, é considerada correta. A tabela 7 abaixo mostra os resultados da ELM usando essa métrica.

Tabela 7: Acurácia folgada dos conjuntos de treino final e de teste.

Acurácia folgada	
Treino final	96.8 %
Teste	93.5 %

Percebe-se que os resultados nesse caso são extraordinários. Tanto o treino, quanto o teste apresentaram acurácia folgada acima de 90 %. Isso indica que quando o classificador erra, ele provavelmente está dando somente um ponto a mais ou um ponto a menos do que o valor registrado como a qualidade da amostra. Pensando em uma aplicação real, onde se quer separar os vinhos entre ruins, médios e bons, pode-se dizer que esse classificador seria de grande ajuda.

e)

Nessa etapa, usou-se a análise por discriminante linear de Fisher (LDA) para realizar a classificação. O LDA determina fronteiras lineares de decisão no espaço de entrada para classificar as amostras. Como uma ELM treinada para resolver o problema do vinho deu resultados bem medíocres em termos de acurácia, é bem improvável que os dados sejam linearmente separáveis. Portanto, espera-se de antemão que a acurácia do LDA seja baixa.

Não há nenhum hiperparâmetro a ser ajustado no LDA. Assim, pode-se concatenar o conjunto de treino ao de validação em um conjunto de treino final para o treino do LDA. Após realizar o treinamento, a acurácia nos conjuntos de treino final e de teste foram calculadas. Os resultados são mostrados na tabela 8.

Tabela 8: Acurácia dos conjuntos de treino final e do teste para o LDA.

Acurácia	
Treino final	50,1%
Teste	48,9 %

Novamente, observou-se resultados bem medíocres. Os resultados foram um pouco piores do que para a ELM, indicando novamente a má decisão de se usar um classificador linear para esse problema, pensando em termos de acurácia.

Por outro lado, a acurácia folgada apresentou resultados excelentes para o LDA. Isso surpreendeu muito o grupo, porque os valores alcançados foram quase tão bons quanto o da ELM. A matriz de confusão para o conjunto de treino e a acurácia folgada para ambos os conjuntos de dados são mostradas abaixo.

Tabela 9: Matriz de confusão do LDA.

	3	4	5	6	7	8	9
3	1	0	1	2	0	0	0
4	0	0	6	23	4	0	0
5	0	0	142	134	6	0	0
6	0	0	110	288	51	0	0
7	0	0	20	108	48	0	0
8	0	0	6	14	15	0	0
9	0	0	0	0	1	0	0

Tabela 10: Acurácia folgada dos conjuntos de treino final e do teste para o LDA.

Acurácia folgada	
Treino final	92.8 %
Teste	92.1 %

f)

Nesta etapa, o grupo testou um terceiro classificador para estimar a qualidade do vinho. O grupo treinou uma Support Vector Machine (SVM), que é considerado um dos classificadores mais poderosos disponíveis.

Uma vez que a relação entre as classes e os atributos selecionados têm se mostrado bastante não lineares, o grupo optou por utilizar a função kernel RBF (do inglês, *radial basis function*). Embora a função Linear tenha capacidade teórica de lidar com tal situação, esta mostrou-se sistematicamente desvantajosa em testes preliminares. Além disso, segundo Lin et al³, é possível encontrar um kernel RBF com a mesma performance de um dado kernel linear.

Assim, o grupo optou por utilizar o kernel RBF, tendo como hiperparâmetros a penalização C e o raio de influência das amostras γ . O conjunto de validação foi usado para determinar os hiperparâmetros. Para definir os valores adequado de C e γ foi feita uma varredura combinada na qual variaram-se os valores em uma grade logarítmica de 10^{-3} a 10^{+3} .

Por fins de comparação com as demais abordagens, o grupo manteve a acurácia como métrica de desempenho. Os parâmetros ótimos obtidos foram de C de 10, 100 e 1000 e de $\gamma = 10$, conforme tabela 11. Foi escolhido o valor de $C = 10$ para o conjunto de teste, cuja acurácia pode ser vista na tabela 12. Percebe-se prontamente que a SVM foi a máquina apresentando os melhores resultados em termos de acurácia.

Tabela 11: Acurácia do conjunto de validação no treino da SVM.

$C \setminus \gamma$.001	.01	.1	1	10	100	1000
.001	0.444	0.444	0.444	0.444	0.444	0.444	0.444
.01	0.444	0.444	0.444	0.444	0.444	0.444	0.444
.1	0.444	0.454	0.515	0.495	0.444	0.444	0.444
1	0.457	0.522	0.537	0.554	0.591	0.573	0.571
10	0.514	0.518	0.533	0.560	0.596	0.574	0.573
100	0.523	0.538	0.527	0.573	0.596	0.574	0.573
1000	0.519	0.542	0.536	0.568	0.596	0.574	0.573

³ Keerthi, S. Sathiya, and Chih-Jen Lin. "Asymptotic behaviors of support vector machines with Gaussian kernel." *Neural computation* 15.7 (2003): 1667-1689.

Tabela 12: Acurácia dos conjuntos de treino final e do teste para o SVM.

Acurácia	
Treino final	59,6%
Teste	58,4 %

Novamente, assim como no caso da ELM, para observar nosso classificador, foi calculada sua matriz de confusão. Ela é mostrada na tabela 13 e, novamente, observa-se a maior parte dos erros a uma unidade de distância do valor considerado correto. Além disso, nota-se que a maior parte dos erros foram obtidos na classe 6, o que, de certa forma, indica que o classificador resultante do treinamento, tende a apostar mais nesta classe, o que, de certa forma, condiz com o esperado, por esta classe ser a mais ocorrente.

Tabela 13: Matriz de confusão do SVM

	3	4	5	6	7	8	9
3	0	0	0	1	0	0	0
4	0	1	1	21	0	0	0
5	1	0	91	134	4	0	0
6	0	3	22	313	9	1	0
7	0	0	2	96	51	2	0
8	0	0	0	19	0	11	0
9	0	0	0	1	0	0	0

A tabela 15 abaixo mostra os resultados da SVM usando como métrica o conceito de acurácia folgada anteriormente discutido. Os parâmetros ótimos obtidos foram de $C = 1000$ e de $\gamma = 0.001$, conforme podemos ver na tabela 14.

Tabela 14: Acurácia folgada do conjunto de validação no treino da SVM.

$C \setminus \gamma$.001	.01	.1	1	10	100	1000
.001	0.930	0.930	0.930	0.930	0.930	0.930	0.930
.01	0.930	0.930	0.930	0.930	0.930	0.930	0.930
.1	0.930	0.936	0.941	0.934	0.930	0.930	0.930
1	0.936	0.945	0.946	0.944	0.943	0.945	0.945
10	0.944	0.952	0.945	0.918	0.932	0.945	0.945
100	0.944	0.949	0.946	0.903	0.932	0.945	0.945
1000	0.953	0.945	0.940	0.904	0.932	0.945	0.945

Tabela 15: Acurácia folgada dos conjuntos de treino final e do teste para o SVM.

Acurácia folgada	
Treino final	95,3 %
Teste	94,0 %

Percebe-se que, também para o SVM utilizado como classificador, tanto o treino, quanto o teste apresentaram acurácia folgada acima de 90 %, o que era esperado de uma máquina de vetor suporte.

g)

Podemos observar pela tabela 16, que os resultados de acurácia obtidos pelas três abordagens estudadas foram bastante sofríveis, embora condizentes com a literatura⁴. O uso da SVM como classificador obteve o melhor resultado. Para o problema de classificar a nota correta de vinho, sem dúvidas a melhor escolha é a SVM.

Tabela 16: Comparação das acurácias obtidas pelos classificadores avaliados: ELM, LDA e SVM.

Acurácia	Treino final	Teste
ELM	54,7 %	52,3 %
LDA	50,1%	48,9 %
SVM	59,6%	58,4 %

Pela tabela 17, quando modificamos a métrica de modo que esta esteja mais apropriada para o problema de classificação, obtemos uma grande melhora nos desempenhos de todas as máquinas estudadas, ficando estas com valores de acurácia acima de 90%.

⁴ <https://onlinecourses.science.psu.edu/stat857/node/229>

Tabela 17: Comparação das acurácias folgadas obtidas pelos classificadores avaliados: ELM, LDA e SVM.

Acurácia folgada	Treino final	Teste
ELM	96.8 %	93.5 %
LDA	92.8 %	92.1 %
SVM	95,3 %	94,0 %

Nessa situação, recomenda-se tanto o uso da SVM quanto da LDA. A SVM apresenta a melhor acurácia no conjunto de teste e o LDA é o modelo mais fácil e simples de ser implementado. Portanto, a escolha acaba dependendo do compromisso complexidade/resultados corretos que o usuário está disposto a fazer. A ELM não se mostrou muito interessante nesse problema, porque ela não é tão simples quanto o LDA, nem fornece resultados tão bons quanto a SVM.

Para um melhor desempenho da ELM, pode-se fazer uso de técnicas mais sofisticadas de regularização, como: ridge regression, LASSO e Elastic Net. Porém, isto se daria ao custo de mais um hiperparâmetro para ser otimizado⁵.

Vale mencionar a observação dos dados nas matrizes de confusão (tabelas 9 e 11). A maior parte dos erros cometidos ocorreu quando as máquinas classificaram como vinho classe 6, vinhos que eram de outra classe. Isso indica que o classificador resultante do treinamento, tende a apostar mais nesta classe, o que, de certa forma, condiz com o esperado, por esta classe ser a mais ocorrente.

Como proposta de melhora nesse problema na métrica acurácia, poderia-se pegar mais amostras de outras classes para balancear as saídas. Outra proposta interessante seria gerar dados de forma sintética para resolver esse mesmo problema. Valores maiores de acurácia também podem ser alcançados com o uso de ensembles.

Os códigos elaborados para esta lista, bem como este relatório com as imagens em cores, podem ser acessados em (<https://github.com/svbueno/NeuralNetwork/tree/master/II/List2>).

⁵ Kulaif, A.C.P. “Técnicas de regularização para máquinas de aprendizado extremo”, Dissertação de Mestrado, Faculdade de Engenharia Elétrica e de Computação, Unicamp, 2014.

Anexo

1. a) $x_{+1} \rightarrow$ escolher classe +1
 $x_{-1} \rightarrow$ escolher classe -1

$$R(x_{+1}/x) = P(C_{+1}/x)$$

$$R(x_{-1}/x) = P(C_{-1}/x)$$

$$P(C_i/x) = \frac{P(x/C_i)P(C_i)}{P(x)}$$

Podemos ignorar o termo $P(x)$ durante a comparação.

$$R(x_i/x) = P(x/C_j)P(C_j) \quad j \neq i \quad F_i(x) = R(x_i/x)$$

Vetor de entrada x

$$F_{-1}(x) = \frac{P(C_{+1})}{\sqrt{2\pi}} \exp\left[-\frac{(x-2)^2}{2}\right]$$

$$F_{+1}(x) = \begin{cases} P(C_{-1})P(x/C_{-1}) & \\ 0,6, & x \in [0,1] \\ 0, & \text{c.c.} \end{cases}$$

$$i = \arg \min_i F_i(x)$$

Decisão
 $\min_i F_i(x)$
 i

$$1. b) P(\text{erro}) = 1 - P(\text{acerto})$$

$$P(\text{acerto}) = P(C_{+1}) P(X \in R_1 / C_{+1}) + P(C_{-1}) P(X \in R_{-1} / C_{-1})$$

$$= 0,4 \int_{R_1} p_G(x/C_{+1}) dx + 0,6 \cdot \int_{R_2} p_G(x/C_{-1}) dx =$$

$$0,6 + 0,4 \left[\int_{-\infty}^0 p_G dx + \int_1^{\infty} p_G dx \right]; X|C_{+1} \sim N(2,1)$$

$$\int_a^b p_G(x/C_{+1}) dx = P(a \leq X \leq b) =$$

$$P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

onde Φ é a cdf da distribuição $N(0,1)$

$$\text{Logo, } \int_{-\infty}^0 p_G dx = \Phi(-2) =$$

$$\text{e } \int_1^{\infty} p_G dx = 1 - \Phi(1)$$

$$P(\text{acerto}) = 0,6 + 0,4 + 0,4 (\Phi(-2) - \Phi(-1)) =$$

$$1 - 0,4 \cdot 0,13586 = 0,946 = 94,6\%$$

$\Phi(-2), \Phi(-1)$ foram calculados com uma tabela normal padrão.

$$P(\text{erro}) = 5,4\%$$