2o Teste de Aprendizagem Autom´atica

3 p´aginas de enunciado com 6 perguntas mais 2 folhas de resposta. Dura¸c˜ao: 1h 30m DI, FCT/UNL, 21 de Dezembro de 2017

Pergunta 1 [4 valores] Considere um problema de classifica¸c˜ao

com classes de hip´oteses *H* com um n´umero infinito de hip´oteses. Podemos estimar um limite superior para o erro verdadeiro da hip´otese com menor erro emp´ırico *h*ˆ em rela¸c˜ao ao seu erro emp´ırico *E*ˆ(*h*ˆ). H´a apenas uma probabilidade de *δ* do erro ver dadeiro ultrapassar esta soma, onde *V C* ´e a dimens˜ao Vap nik–Chervonenkis do classificador e *m* o tamanho do conjunto de treino:

Dimens˜ao Bias Var 2 0.308 0.001 3 0.252 0.002 4 0.104 0.005 5 0.012 0.102 6 0.008 0.120 7 0.006 0.328

*E*ˆ(*h*ˆ) + *O*

s

*V C*(*H*)

*m*ln *m*

*V C*(*H*)+1*m*ln 1*δ*

!

Partindo de um conjunto de dados a duas dimens˜oes, a tabela mostra os valores de *bias* e *variance* calculados com classificadores lineares aplicados aos dados originais (dimens˜ao 2) e a expans˜oes n˜ao lineares dos dados originais para mais dimens˜oes (3 a 7).

1.a) Indique a dimens˜ao de um destes classificadores em que o termo *E*ˆ(*h*ˆ) seja o mais importante na estimativa do limite superior do erro verdadeiro. Justifique a sua resposta.

q*V C*(*H*)

*m*ln *m*

1.b) Indique a dimens˜ao de um destes classificadores em que o termo

*V C*(*H*) +1*m*ln 1*δ*seja

o mais importante na estimativa do limite superior do erro verdadeiro. Justifique a sua resposta.

Pergunta 2 [4 valores] Num problema de classifica¸c˜ao com duas classes queremos reduzir a dimens˜ao de quatro para dois atributos. O painel da esquerda da figura abaixo mostra os gr´aficos de kernel density estimation da distribui¸c˜ao dos valores de cada atributo (X1 a X4), na mesma escala, com as distribui¸c˜oes para os exemplos da classe 1 na linha de cima e as distribui¸c˜oes dos exemplos da classe 2 na linha de baixo. O painel da direita mostra a matriz de gr´aficos de dispers˜ao (*scatter matrix*) para estes dados.



2.a) Quais os dois atributos que seleccionaria, destes quatro? Justifique a sua resposta e se houver alternativas enumere-as.

2

2.b) Suponha que em vez de seleccionar dois atributos fez uma an´alise de componentes principais (*Principal Component Analysis*) e projectou os dados nos dois componentes com maior valor pr´oprio, com o principal no eixo horizontal o segundo no eixo vertical. Indique qual dos trˆes gr´aficos abaixo (A, B, C) corresponde a esta projec¸c˜ao explicando porque ´e que os outros dois gr´aficos n˜ao podem ser os correctos. Note que as escalas s˜ao iguais nos 3 gr´aficos.



Pergunta 3 [3 valores]

A figura `a direita mostra o resultado das duas pri

meiras itera¸c˜oes de um algoritmo de clustering. A 

primeira itera¸c˜ao separou os exemplos em dois gru

pos indicados pela cor, com os triˆangulos a negro

num grupo e os restantes, a cinzento, no outro. A

segunda itera¸c˜ao criou dois clusters adicionais com

os exemplos do grupo a cinzento, diferenciados pelos

quadrados e pelos c´ırculos. Para cada um destes trˆes

algoritmos, indique se pode ou n˜ao pode ter sido esse

o algoritmo usado, justificando a sua resposta:

3.a) Fuzzy C-Means

3.b) Bissecting K-Means

3.c) Agglomerative Clustering

Pergunta 4 [6 valores] O mesmo conjunto de dados bidimensionais foi agrupado com algoritmos de *clustering* diferentes. Em todos os casos foi usada a distˆancia euclideana como medida de distˆancia entre pontos. Leia as quatro al´ıneas desta pergunta primeiro e observe atentamente os gr´aficos que se seguem antes de responder justificando cada resposta. No gr´afico, cada s´ımbolo maior representa um ponto num *cluster*. Os pontos que n˜ao foram atribu´ıdos a um *cluster* s˜ao representados com c´ırculos pequenos (gr´aficos A e D).

4.a) Qual gr´afico (A, B, C, D) mostra o resultado da aplica¸c˜ao do algoritmo Density-based spatial clustering of applications with noise (DBSCAN), com um valor de 10 para o n´umero m´ınimo de vizinhos para os pontos de *core*?

4.b) Qual gr´afico mostra o resultado da aplica¸c˜ao do algoritmo DBSCAN com um valor de 5 para o n´umero m´ınimo de vizinhos para os pontos de *core*?

4.c) Qual gr´afico mostra o resultado de *clustering* usando o algoritmo de k-means com um valor de *k* = 2?

4.d) Qual gr´afico mostra o resultado de *clustering* usando uma mistura de duas distribui¸c˜oes Gaus sianas?

3

A B C D

Pergunta 5 [2 valores] O treino de modelos ocultos de Markov (*Hidden Markov Models*) e de algoritmos de clustering como K-Means e misturas de gaussianas (*Gaussian Mixture Models*) exige resolver o problema de maximizar a verosimilhan¸ca de um conjunto de parˆametros sem se saber todas as vari´aveis a modelar (por exemplo, os estado ocultos no HMM ou as atribui¸c˜oes aos clusters no K-Means e GMM). Explique sucintamente como se resolve esse problema.

Pergunta 6 [1 valores] Em teoria, uma rede neuronal com uma camada oculta pode aproximar qualquer fun¸c˜ao simplesmente aumentando o n´umero de neur´onios na camada oculta. Explique porque ´e que, ainda assim, ´e geralmente prefer´ıvel usar uma rede neuronal profundas em vez de uma rede neuronal com uma camada oculta suficientemente grande.

1a) 1b)

2a) 2b)

3a) 3b) 3c)

4a) 4b) 4c) 4d)

5)

6)