1o Teste de Aprendizagem Autom´atica

2 p´aginas com 9 perguntas mais 2 folhas de resposta. Dura¸c˜ao: 1 hora e 30 minutos DI, FCT/UNL, 25 de Outubro de 2017

Pergunta 1 [1.5 valores] Explique porque ´e que se costuma minimizar o erro quadr´atico m´edio para treinar modelos em problemas de regress˜ao.

Pergunta 2 [2.5 valores]

Separou-se, aleatoriamente, um conjunto de 150 pontos bidi

mensionais (x e y) em 3 conjuntos de 50 pontos cada (conjuntos A, B e C). A tabela `a direita mostra o erro quadr´atico m´edio da previs˜ao do y em fun¸c˜ao do x, para cada conjunto, usando polin´omios de grau 2 a 7 cujos coeficientes foram obtidos mini mizando o erro quadr´atico m´edio medido no conjunto A.

Indique o melhor grau para o polin´omio a usar nesta regress˜ao e indique uma estimativa n˜ao tendenciosa do erro verdadeiro dessa hip´otese. Se houver alternativas, indique todas. Justifique a sua resposta.

Grau Conj. A Conj B Conj C 2 3.8 4.2 3.9 3 3.1 3.5 3.3 4 1.2 1.6 2.1 5 0.9 1.9 2.0 6 0.6 6.2 7.7 7 0.1 6.6 9.6

Pergunta 3 [2 valores] O erro de valida¸c˜ao cruzada d´a uma estimativa do erro do modelo ou da hip´otese? Justifique a sua resposta.

Pergunta 4 [2 valores] Explique o que ´e a regulariza¸c˜ao e descreva um exemplo de como se usa a regulariza¸c˜ao num destes casos (`a sua escolha): regress˜ao polinomial, regress˜ao log´ıstica ou redes neuronais.

Pergunta 5 [1 valores] Indique como desenhar uma rede neuronal de forma a poder separar classes de exemplos que n˜ao sejam linearmente separ´aveis e explique como ´e que isso permite tal separa¸c˜ao.

Pergunta 6 [4 valores] Para criar um classificador, estimou-se a probabilidade conjunta da classe e dos valores dos atributos multiplicando a probabilidade a priori de um ponto pertencer a cada classe (a propor¸c˜ao da classe no conjunto de treino) pela probabilidade condicional do valor de cada atributo nessa classe. Assim, para classificar um ponto, procura-se a classe que maximiza a probabilidade conjunta de acordo com esta express˜ao:

argmax

*k∈{*0*,*1*,...,K}*

ln *p*(*Ck*) +X*N j*=1

ln *p*(*xj |Ck*)

6.a) Explique o que estamos a assumir acerca das distribui¸c˜oes de probabilidades dos atributos para que a express˜ao da probabilidade conjunta seja assim t˜ao simples.

6.b) Se usarmos Kernel Density Estimation para estimar os logaritmos das probabilidades condicionais dos atributos em cada classe (*lnp*(*xj |Ck*)) podemos considerar este classificador um exemplo de *lazy learning*? Justifique a sua resposta.

2

Pergunta 7 [4 valores]

A figura `a direita mostra os dados, a linha discriminante entre as duas classes (linha cont´ınua) e as margens (tra cejado) para um classificador que prevˆe a classe de cada exemplo *~xt* pelo sinal da seguinte express˜ao: 

X*N n*=1

*αnynK*( *~xn, ~xt*)

Os valores de *αn* para os vectores do conjunto de treino foram calculados pela solu¸c˜ao desta minimiza¸c˜ao:

*minα*

1

2

X*N n*=1

X*N m*=1

*αnαmynymK*( *~xm, ~xn*) *−*X*N n*=1

!

*αn*

sujeita ´as seguintes restri¸c˜oes:

0 *≤ αn ≤* 5*, n* = 1*, ..., N* X*N*

*n*=1

A fun¸c˜ao *K*( *~xm, ~xn*) foi uma destas duas:

*•* Op¸c˜ao A: *K*( *~xm, ~xn*) = *~xmT~xn*

*•* Op¸c˜ao B: *K*( *~xm, ~xn*) = ( *~xmT~xn*)3

*αnyn* = 0

7.a) Qual das duas op¸c˜oes (A ou B) foi usada para a fun¸c˜ao *K*( *~xm, ~xn*)? Justifique a as resposta.

7.b) Indique os valores dos multiplicadores *α* para os vectores X, Y e Z indicados na imagem. Se n˜ao puder determinar o valor exacto do *α* em algum destes casos, indique o intervalo de valores poss´ıveis para esse vector.

Pergunta 8 [1.5 valores] Explique porque ´e que n˜ao ´e ´util usar o m´etodo de Bootstrap Aggregating (Bagging) com modelos cujo *bias* seja maior do que a variˆancia.

Pergunta 9 [1.5 valores] No algoritmo AdaBoost, treina-se sempre o mesmo modelo a cada itera¸c˜ao, com os mesmos dados. No entanto, a cada itera¸c˜ao obt´em-se hip´oteses diferentes. Explique porquˆe.

1) 2)

3) 4)

5)

6 a) 6 b)

7 a)

7 b) 8)

9)