

1º Teste de Aprendizagem Automática

2 páginas com 8 perguntas e 2 folhas de resposta. Duração: 1h 30m
DI, FCT/UNL, 9 de Novembro de 2016

Pergunta 1 [4 valores] O André tem um conjunto de 50 pontos, com os valores x e y de cada ponto, e quer fazer uma regressão para poder prever os valores y de pontos futuros a partir dos valores de x . Para isso, usou vários modelos, treinando cada um no conjunto de 50 pontos e medindo o erro quadrático médio nesses 50 pontos. Os resultados do André foram:

Modelo	Erro quadrático médio
1: $y = \theta_1 x + \theta_0$	1.3
2: $y = \theta_2 x^2 + \theta_1 x + \theta_0$	0.8
3: $y = \theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0$	0.2
4: $y = \theta_4 x^4 + \theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0$	0.05

1.a) O André ficou muito feliz com o modelo 4, porque tinha um erro menor, e disse à Beatriz que este era o melhor modelo. O André tem razão? Justifique a sua resposta.

A Beatriz dividiu os pontos do André em dois conjuntos de 25 pontos, seleccionados aleatoriamente. Depois, criou 100 réplicas do primeiro conjunto de 25 pontos. Cada réplica consiste em 25 pontos retirados aleatoriamente, com reposição, do conjunto inicial. Para cada réplica, a Beatriz treinou cada um dos modelos com os pontos na réplica e calculou estas duas medidas para o outro conjunto de 25 pontos que deixou de fora:

Medida	Fórmula
A: erro da previsão média	$\frac{1}{N} \sum_{n=1}^N (\bar{y}(x_n) - t_n)^2$
B: dispersão das previsões	$\frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M (\bar{y}(x_n) - y_m(x_n))^2$

(y é a previsão; t o valor real; N o número de pontos e M o número de réplicas)

Os valores destas medidas da Beatriz, para cada um dos modelos, foram:

Modelo	Medida A	Medida B
1: $y = \theta_1 x + \theta_0$	1.4	0.1
2: $y = \theta_2 x^2 + \theta_1 x + \theta_0$	0.8	0.2
3: $y = \theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0$	0.2	0.25
4: $y = \theta_4 x^4 + \theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0$	0.03	1.7

1.b) Com base nos resultados da Beatriz, indique qual o melhor modelo dos quatro que o André usou e explique como chegou a essa conclusão.

1.c) De entre os modelos do André, consegue dar um exemplo de um modelo que esteja em *underfitting* e um que esteja em *overfitting*? Justifique a sua resposta.

Pergunta 2 [3 valores] Com *Logistic Regression* faz-se uma regressão para aproximar a função que dá a probabilidade de cada ponto pertencer a uma de duas classes. Essa função depende dos valores dos atributos \vec{x} e dos parâmetros a ajustar \tilde{w} :

$$g(\vec{x}, \tilde{w}) = P(C_1|\vec{x}) = 1 - P(C_0|\vec{x}) = \frac{1}{1 + e^{-(\tilde{w}^T \vec{x} + w_0)}}$$

2.a) Explique como é que se pode usar esta regressão para classificar pontos em duas classes (nota: não precisa de fazer cálculos ou demonstrações; apenas de explicar a ideia).

2.b) Se estas classes, nestes pontos, não forem linearmente separáveis, explique como pode transformar o conjunto de pontos de forma a conseguir separá-los com *Logistic Regression*.

Pergunta 3 [3 valores] Foi criado um classificador de *Naïve Bayes* usando um conjunto de treino com duas classes, A e B, em que cada ponto tem dois atributos contínuos e as classes estão representadas na mesma proporção no conjunto de treino, 50% cada uma. Modelou-se a distribuição de probabilidade de cada atributo, em cada classe, usando *Kernel Density Estimation*.

3.a) Explique como se pode usar estas distribuições de probabilidade para prever a classe de um ponto a partir dos valores dos atributos.

3.b) Explique como se pode usar estas distribuições de probabilidade para gerar pontos artificiais semelhantes aos pontos de uma dada classe, A ou B.

Pergunta 4 [3 valores] Considere a técnica de validação cruzada (*Cross Validation*).

4.a) A validação cruzada dá-nos uma estimativa do erro verdadeiro de um modelo ou de uma hipótese? Justifique sucintamente a sua resposta.

4.b) Descreva sucintamente o que se tem de fazer para calcular o erro de validação cruzada.

4.c) Se usarmos o erro de validação cruzada para otimizar um parâmetro de regularização, o erro calculado será um estimador tendencioso ou não tendencioso do erro verdadeiro? Justifique a sua resposta.

Pergunta 5 [3 valores] A máquina de vector de support (*Support Vector Machine*, SVM) é treinada obtendo os valores de α que maximizam a expressão:

$$\sum_{n=1}^N \alpha_n - \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m K(\vec{x}_m, \vec{x}_n)$$

onde N é o número de exemplos, y o valor da classe de cada exemplo (representados com um círculo preenchido a cinzento para a classe 1 e um círculo preenchido a branco para a classe -1) e \vec{x} o vector com as coordenadas de cada ponto. Adicionalmente, impõem-se as seguintes restrições:

$$\alpha_n \geq 0, \quad n = 1, \dots, N \quad \sum_{n=1}^N \alpha_n y_n = 0$$

5.a) Indique como podemos identificar os vectores de suporte depois de treinar a SVM.

5.b) Indique como se pode regularizar o treino da SVM, permitindo que os pontos penetrem as margens ou mesmo sejam mal classificados.

5.c) Explique o que é a função $K(\vec{x}_m, \vec{x}_n)$ e o papel que esta tem no chamado *kernel trick*.

Pergunta 6 [2 valores] Um perceptrão de múltiplas camadas (*Multilayer Perceptron*, MLP) pode classificar correctamente conjuntos de pontos nos quais as classes não sejam linearmente separáveis. No entanto, se só tiver uma camada, isto não é possível. Explique o que acontece na camada escondida do MLP que permite a classificação de pontos cujas classes não são linearmente separáveis.

Pergunta 7 [1 valores] O método de *bootstrap aggregation* (*bagging*) de classificadores consiste em treinar um modelo em muitas réplicas do conjunto de treino, gerando muitas hipóteses, e classificar os pontos pela maioria dos classificadores. Este método é mais eficaz com classificadores instáveis, que variem muito com o conjunto de treino, ou com classificadores robustos, mais resistentes a variações nos pontos usados para os treinar? Justifique a sua resposta.

Pergunta 8 [1 valores] Descreva o resultado obtido no final do treino com o método de *Adaptive Boosting* (*AdaBoost*) e explique como se usa esse resultado para prever a classe de exemplos de classe desconhecida.

AA Teste 1, 9-11-2016

Numero: _____

Preencha o seu nome abaixo e o seu número à direita. Pinte por baixo de cada dígito do seu número o círculo correspondente. Por fim indique o número de filas de alunos à sua frente e o número de alunos à sua direita pintando o círculo correspondente abaixo.

Nome:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Filas à Frente	<input type="radio"/>															
Alunos à Direita	<input type="radio"/>															

0	<input type="radio"/>				
1	<input type="radio"/>				
2	<input type="radio"/>				
3	<input type="radio"/>				
4	<input type="radio"/>				
5	<input type="radio"/>				
6	<input type="radio"/>				
7	<input type="radio"/>				
8	<input type="radio"/>				
9	<input type="radio"/>				

1a)

1b)

1c)

2a)

2b)

3a)

3b)

AA Teste 1, 9-11-2016

Numero: _____

Preencha o seu nome abaixo e o seu número à direita. Pinte por baixo de cada dígito do seu número o círculo correspondente. Por fim indique o número de filas de alunos à sua frente e o número de alunos à sua direita pintando o círculo correspondente abaixo.

Nome:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Filas à Frente	<input type="radio"/>															
Alunos à Direita	<input type="radio"/>															

0	<input type="radio"/>				
1	<input type="radio"/>				
2	<input type="radio"/>				
3	<input type="radio"/>				
4	<input type="radio"/>				
5	<input type="radio"/>				
6	<input type="radio"/>				
7	<input type="radio"/>				
8	<input type="radio"/>				
9	<input type="radio"/>				

4a)

4b)

4c)

5a)

5b)

5c)

6)

7)

8)