

Exame de Aprendizagem Automática

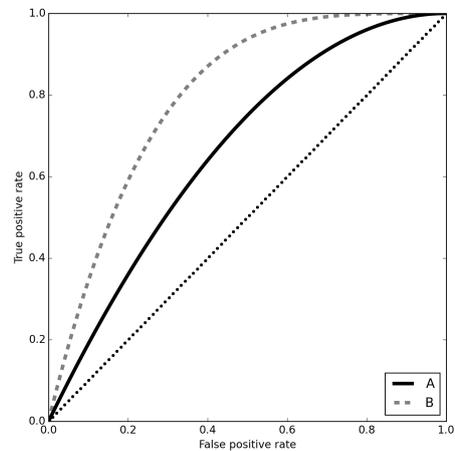
2 páginas com 11 perguntas e 3 folhas de resposta. Duração: 2 horas e 30 minutos
DI, FCT/UNL, 5 de Janeiro de 2016

Nota: O exame está cotado para 40 valores. Os 20 valores das perguntas de 1 a 6 correspondem ao primeiro teste; os restantes ao segundo teste. A nota da componente teórico-prática será a média destas duas partes, usando para cada uma a melhor nota entre o exame e o teste correspondente.

Pergunta 1 [4 valores] Suponha que quer seleccionar o melhor grau do polinómio a usar num problema de regressão. Usa um terço dos dados disponíveis para treinar os vários polinómios, um terço para validação, e um terço para teste do polinómio escolhido no final. Explique porque é que necessita deste conjunto final de teste, se já tinha calculado o erro de validação para cada polinómio.

Pergunta 2 [3 valores] Numa regressão polinomial, a linha é colocada de forma a minimizar o erro quadrático entre os valores conhecidos e os valores previstos. Na regressão logística, que apesar do nome é um classificador, onde é que se quer colocar o hiperplano que separa as duas classes?

Pergunta 3 [3 valores] A figura à direita mostra as curvas ROC (*Receiver Operating Characteristic*) para dois classificadores. Ambos foram treinados com o mesmo conjunto de treino, e ambas as curvas foram calculadas com o mesmo conjunto de teste. O eixo vertical mostra a fracção de verdadeiros positivos, o horizontal a fracção de falsos positivos. Indique, justificando, qual dos dois classificadores (A ou B) é o melhor.



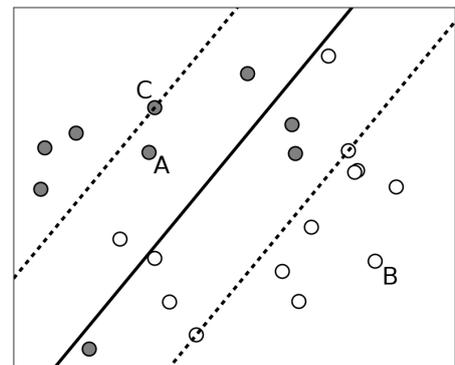
Pergunta 4 [2 valores] Descreva como se treina um classificador K-Nearest Neighbours.

Pergunta 5 [4 valores] A figura à direita mostra esta *Support Vector Machine*:

$$\min_{\alpha} \left(\frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m \vec{x}_m^T \vec{x}_n - \sum_{n=1}^N \alpha_n \right)$$

Foram impostas as seguintes restrições durante a minimização:

$$0 \leq \alpha_n \leq 10, \quad n = 1, \dots, N \quad \sum_{n=1}^N \alpha_n y_n = 0$$



Indique, **justificando**, qual o valor ou valores possíveis de α para cada um dos pontos A, B e C indicados na figura. Se quiser indicar um intervalo, pode usar a notação $[a, b]$ para um intervalo de valores entre a e b incluindo a e b , $]a, b[$ para um intervalo de valores entre a e b excluindo a e b , ou $]a, b]$, $[a, b[$ para excluir apenas um dos extremos.

Pergunta 6 [4 valores] O classificador *Naïve Bayes Classifier* assume que todos os atributos são condicionalmente independentes dada a classe. Por outro lado, o classificador de Bayes completo (*Bayes Classifier*) não assume que os atributos sejam independentes. Sabendo que muitas vezes os atributos não são independentes, mesmo dada a classe, porque é que quase sempre se prefere usar o classificador *Naïve Bayes*?

Pergunta 7 [3 valores] A inequação abaixo estima o limite superior para o erro verdadeiro E da hipótese \hat{h} que minimiza o erro de treino \hat{E} . Este limite é função de δ , do tamanho do conjunto de treino m e da dimensão Vapnik-Chervonenkis da classe de hipóteses, $VC(\mathcal{H})$, e limita o erro verdadeiro com probabilidade $1 - \delta$.

$$E(\hat{h}) \leq \hat{E}(\hat{h}) + \mathcal{O} \left(\sqrt{\frac{VC(\mathcal{H})}{m} \ln \frac{m}{VC(\mathcal{H})} + \frac{1}{m} \ln \frac{1}{\delta}} \right)$$

Recorrendo a esta inequação, explique porque é que aumentar o número de atributos (*features*) num problema de classificação com um classificador linear aumenta o risco de sobreajustamento (*overfitting*).

Pergunta 8 [2 valores] Explique um método para visualizar conjuntos de dados com mais de 3 atributos. Assuma que todos os atributos são numéricos e de valores contínuos.

Pergunta 9 [2 valores] Explique um método para reduzir o número de atributos de um conjunto de dados.

Pergunta 10 [10 valores] Indique, **justificando a resposta**, o algoritmo de *clustering* e os parâmetros mais relevantes que utilizaria nos seguintes casos:

10.a) Tem um conjunto com 200 pontos que quer reduzir a 20 pontos representativos da distribuição do conjunto inicial.

10.b) Tem um conjunto com 200 pontos e razões para crer que cada um dos pontos foi gerado por um de três processos aleatórios diferentes, cada um com a sua distribuição de probabilidades com pontos médios e variâncias diferentes.

10.c) Tem um conjunto com 200 pontos, quer atribuir todos os pontos a algum *cluster*, mas não faz ideia de quantos *clusters* deverá criar.

10.d) Tem um conjunto com 200 pontos que quer agrupar, não sabe quantos *clusters* devem ser gerados, mas não quer que os pontos isolados sejam atribuídos a algum *cluster*.

10.e) Tem um conjunto com 200 pontos que quer agrupar em 12 *clusters*, mas quer perceber também as relações entre esses *clusters* para saber quais se agrupam em grupos maiores, numa árvore com todos os 12 *clusters*.

Pergunta 11 [3 valores] Considere o problema de treinar redes neuronais artificiais profundas (com várias camadas escondidas).

11.a) Explique a razão de ser mais difícil treinar redes neuronais artificiais profundas do que treinar redes com apenas uma camada oculta.

11.b) Explique uma forma de se ultrapassar essa dificuldade.

AA Exame de Recurso, 2016-01-05

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"/>				

1)

2)

3)

4)

5)

AA Exame de Recurso, 2016-01-05

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"/>				

6)

Segunda Parte

7)

8)

9)

AA Exame de Recurso, 2016-01-05



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"/>				

10 a)

10 b)

10 c)

10 d)

10 e)

11 a)

11 b)