

Exame de Aprendizagem Automática

2 páginas com 12 perguntas e 3 folhas de resposta. Duração: 2 horas e 30 minutos

DI, FCT/UNL, 12 de Janeiro de 2017

Nota: O exame está cotado para 40 valores. Os 20 valores das perguntas de 1 a 7 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 cada teste e a parte correspondente do exame.

Primeira Parte (correspondente ao primeiro teste)

Pergunta 1 [4 valores] Suponha que quer treinar um classificador com base num conjunto de exemplos previamente classificados. Tem vários modelos que pode usar. Descreva o que deve fazer para seleccionar o melhor modelo, instanciar o modelo seleccionado na melhor hipótese e obter uma estimativa não enviesada do erro verdadeiro dessa hipótese final.

Pergunta 2 [2 valores] Numa regressão polinomial, a melhor hipótese é encontrada minimizando o erro quadrático entre os valores conhecidos e os valores previstos. Explique a razão deste método: porquê minimizar o erro quadrático?

Pergunta 3 [2 valores] Explique porque é que o classificador K-Nearest Neighbours é considerado “preguiçoso” (*lazy*), apontando os aspectos relevantes do algoritmo.

Pergunta 4 [2 valores] Explique porque é que o classificador *Naïve Bayes* é considerado generativo, apontando os aspectos relevantes do algoritmo.

Pergunta 5 [2 valores] Explique o que é a regularização e descreva um exemplo de como se usa a regularização num destes casos (à sua escolha): regressão polinomial, regressão logística, máquina de vectores de suporte, redes neuronais.

Pergunta 6 [2 valores] O classificador de Bayes completo (*Bayes Classifier*) não assume independência condicional dos atributos dada a classe. O que, geralmente, seria mais correcto porque normalmente os atributos não são independentes, mesmo dada a classe. Explique então qual é a vantagem de usar um classificador *Naïve Bayes*, que assume que os atributos são condicionalmente independentes dada a classe.

Pergunta 7 [6 valores] Suponha que tem um conjunto de exemplos com dois atributos contínuos, x_1 e x_2 , classificados em duas classes, 0 e 1, e suponha que estas classes **não** são linearmente separáveis. Explique, para cada tipo de classificador abaixo, como se resolve o problema das classes não serem linearmente separáveis:

7.a) Regressão logística (*logistic regression*).

7.b) Máquina de vectores de suporte (*support vector machine*).

7.c) Rede neuronal.

Pergunta 8 [4 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} em função de δ , do tamanho do conjunto de treino m e da dimensão Vapnik-Chervonenkis da classe de hipóteses, $VC(\mathcal{H})$, 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:

8.a) Porque é que aumentar o poder para classificar correctamente os exemplos do conjunto de treino pode aumentar o erro verdadeiro da hipótese obtida.

8.b) Porque é que usar modelos demasiado simples pode aumentar o erro verdadeiro da hipótese obtida.

8.c) Porque é que há vantagem em usar conjuntos de treino maiores.

Pergunta 9 [4 valores] Suponha que tem um conjunto de exemplos em que cada exemplo é descrito por 5 atributos numéricos contínuos.

9.a) Explique um método para visualizar este conjunto de exemplos.

9.b) Explique uma forma de seleccionar os 2 melhores atributos para treinar um classificador assumindo que sabe também a classe de cada exemplo.

Pergunta 10 [8 valores] Indique, **justificando a resposta**, o algoritmo de *clustering* e, caso seja relevante, os parâmetros que utilizaria nos seguintes casos:

10.a) Tem um conjunto de 100 exemplos que quer organizar numa árvore binária com 4 níveis. Em cada nível excepto o último, cada grupo tem dois sub-grupos no nível imediatamente abaixo.

10.b) Tem um conjunto com 300 exemplos, cada um com dois atributos numéricos contínuos, e quer agrupá-los nas regiões de maior densidade, descartando os pontos isolados em regiões de baixa densidade.

10.c) Tem um conjunto com 200 exemplos, cada um com dois atributos numéricos contínuos, e suspeita que cada ponto surgiu de uma de quatro distribuições aleatórias gaussianas, cada uma com a sua média e matriz de covariância.

10.d) Tem um conjunto com 500 exemplos obtidos de uma distribuição tridimensional contínua e quer representar essa distribuição com apenas 20 combinações de valores dos 3 atributos.

Pergunta 11 [2 valores] Alguns algoritmos de *clustering* dependem de uma medida de *linkage*. Explique o que é essa medida e dê um exemplo de uma medida de *linkage*.

Pergunta 12 [2 valores] Explique porque é que é mais difícil treinar redes neuronais artificiais profundas do que redes neuronais artificiais com poucas camadas.

AA Exame de Recurso 2017-01-12

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"/>				

Primeira parte

1)

2)

3)

4)

5)

6)

AA Exame de Recurso 2017-01-12

Numero: _____

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"/>				

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"/>															

7 a)
7 b)
7 c)

Segunda Parte

8 a)
8 b)
8 c)
9 a)

9 b) na próxima folha

AA Exame de Recurso 2017-01-12

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"/>				

9 b)

10 a)

10 b)

10 c)

10 d)

11

12