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] André has a set of 50 points with the x e y values for each. He wants to do a regression to predict the y values of future points from their x values. To this end, he trained several models on the set of 50 points and measured the mean quadratic error on those 50 points. These were his results:

Model	Mean quadratic error
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) André was very happy with model 4, because it had the smallest error, and so he told Beatriz this was the best model. Is André right? Justify your answer.

Beatriz split, at random, the data points into two sets of 25 points. She then created 100 replicas of the first set of 25 points. Each replica consists of 25 points randomly drawn, with reposition, from the initial set. For each replica, Beatriz trained each model with the replica points and computed the following measures using the other set of 25 points she had left out at the start:

Measure	Formula
A: mean prediction error	$\frac{1}{N}\sum_{n=1}^{N}(\bar{y}(x_n) - t_n)^2$
B: dispersion of predictions	$\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}(\bar{y}(x_n) - y_m(x_n))^2$

(y is the predicted value; t the true value; N the number of points and M the number of replicas)

The values measured by Beatriz were the following, for each model:

Model	Measure A	Measure 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) Based on the results Beatriz obtained, choose the best of the four models André used and explain how you reached your conclusion.

1.c) From these four models, can you give an example of a model that is *underfitting* and another that is *overfitting*? Justify your answer.

Pergunta 2 [3 valores] With *Logistic Regression* we do use regression to approximate a function that gives the probability of each point belongint to one of two classes. This function depends on the attribute values $\vec{x}$ and on the parameters to be adjusted, $\widetilde{w}$:

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

2.a) Explain how to use this regression to classify points in two classes (note: you do not need to do computations or demonstrations; you need only to explain the idea).

2.b) If these classes are not linearly separable, explain how the data set can be transformed so as to separate the classes with *Logistic Regression*.

Pergunta 3 [3 valores] ANaïve Bayes classifier was created from a training set with two classes, `A` e `B`, in which each example has two continuous attributes and the classes are equally represented in the same proportion in the data set, 50% each. The probability distribution of each attribute in each class was modelled with *Kernel Density Estimation*.

3.a) Explain how those distributions can be used to predict the class of a point from the values of its attributes.

3.b) Explain how these probability distributions can be used to generate artificial examples similar to the points of one given class, A or B.

Pergunta 4 [3 valores] Consider the *Cross Validation* technique.

4.a) Cross validation gives us an estimate of the true error of a model or of a hypothesis? Justify your answer.

4.b) Describe what you need to do to compute the cross validation error.

4.c) If we use the cross validation error to optimize a regularization parameter, the error we computed will be a biased or an unbiased estimator of the true error? Justify your answer.

Pergunta 5 [3 valores] A *Support Vector Machine*, SVM, is trained computing the α values that maximize the expression:

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

where N is the number of examples, y the class labels (1 or minus 1) e $\vec{x}$ o vector with the coordinates for each point. In addition, the following restrictions were imposed:

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

5.a) Explain how we can identify the support vectors after training the SVM.

5.b) Explain how we can regularize the training of the SVM, allowing points to penetrate the margins or even be misclassified.

5.c) Explain what is the function $K(\vec{x_m}, \vec{x_n})$ e its role in the *kernel trick*.

Pergunta 6 [2 valores] A *Multilayer Perceptron*, MLP, can classify correctly sets of points for which the classes are not linearly separable. However, if there is only one layer, this is not possible. Explain what happens in the hidden layer of the MLP to allow the classification of points whose classes are not linearly separable.

Pergunta 7 [1 valores] The *boostrap aggregation* (*bagging*) method for classifiers consists in training a model in many replicas of the training set, generating many hypothesis, and then classifying points from the majority of the classifiers. Is this method more suited for unstable classifiers, who vary a lot with the training set, for robust classifiers that are more resistant to variations in the points used to train them? Justify your answer.

Pergunta 8 [1 valores] Describe what is the result obtained after training with the *Adaptive Boosting* (*AdaBoost*) method and explain how to use this result to predict the class of examples whose class is unknown.

AA Teste 1, 9-11-2016

Numero:

0	○	○	○	○	○										
1	○	○	○	○	○										
2	○	○	○	○	○										
3	○	○	○	○	○										
4	○	○	○	○	○										
5	○	○	○	○	○										
6	○	○	○	○	○										
7	○	○	○	○	○										
8	○	○	○	○	○										
9	○	○	○	○	○										

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 circulo correspondente abaixo.

Nome:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Filas à Frente	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Alunos à Direita	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

1a)

1b)

1c)

2a)

2b)

3a)

3b)

AA Teste 1, 9-11-2016

Numero:

0	○	○	○	○	○
1	○	○	○	○	○
2	○	○	○	○	○
3	○	○	○	○	○
4	○	○	○	○	○
5	○	○	○	○	○
6	○	○	○	○	○
7	○	○	○	○	○
8	○	○	○	○	○
9	○	○	○	○	○

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 circulo correspondente abaixo.

Nome:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Filas à Frente	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Alunos à Direita	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

4a)

4b)

4c)

5a)

5b)

5c)

6)

7)

8)