Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard

11/07/2013

Answer the following questions identifying the key aspects and try not to exceed the 1.5 page limit per question.

- Use only the 3 sheets provided by the teacher
- **Write your answers on different sheets according to the question**
- Write your name and Student ID on each sheet you turn in
- English is the official language, however Italian is allowed
- Pen and pencil are allowed no other technical mean to support yourself

In case you have special needs (e.g., being graded within a given time) please ***tell it to the teacher!***.

Question 1: Linear Discriminant Analysis (Answer on sheet 1 – 8 points)

a) Describe Linear Discriminant Analysis (LDA), its analytical form, its assumptions, and its advantages wrt plain Linear Regression (on the indicator matrix).

b) Derive the formula for Quadratic Discriminant Analysis (QDA) and discuss how a quadratic classifier can also be implemented by Linear Discriminant Analysis on an extended space.

c) How do LDA and QDA work for multi-class problems? What is their complexity in terms of parameters to be estimated?

Question 2: Maximum Margin Classifiers (Answer on sheet 1 – 8 points)

a) Describe the idea of the Maximum Margin Classifier and its formulation as a constrained optimization problem on the β parameters of the linear classifier both in the case when the data are linearly separable and in the case the data are not linearly separable.

b) How the maximization problem is solved and what is the form of the final solution? What data points are involved in this solution and why the maximum margin classifier is also known as Support Vector Machine classifier? Only the linearly separable case is required here, but feel free to discuss also the non linearly separable one.

c) Are Support Vector Machines limited to linear boundaries? How is it possible to extend them to use non linear separating boundaries? What is the so called “kernel trick”?

Question 3: Clustering (Answer on sheet 2 – 8 points)

a) Hierarchical clustering is not a single algorithm but rather a family of different clustering algorithms. Explain (1) how this family is composed, (2) how these algorithms work, and (3) what metrics exist to measure the

distance between clusters. Finally (4), when is a hierarchical algorithm preferrable with respect to another one such as K-Means and when is it not?

b) Some algorithms (e.g. k-means/medoids, fuzzy c-means, spectral clustering) need the number of clusters to be provided in advance. What methods would you use to choose the best number of clusters? Motivate your answer.

c) Among the different algorithms we have studied, there are two which especially rely on the concept of neighbor. Which are these two algorithms and what are their main differences?

Question 4: Regression (Answer on sheet 3 – 8 points)

Let β be the least squares estimate for β , for the linear regression model

$$f(x)=\sum_{i=j}^{n} \beta x_j + \beta_0 \epsilon$$

Let $x=(x_1,\cdots,x_m)$, and let D be a training set of n points $x^{(1)},x^{(2)},\ldots,x^{(n)}$

a) How can you obtain β given D? State the regression problem, formulate the least squares estimator of the regression problem as the solution of an optimization problem, and derive mathematically the formulation for β (include all mathematical derivations)

b) What is the MSE(β)? Do you expect a learning algorithm will minimize or maximize such quantity? Show, by including mathematical derivations, how the MSE can be decomposed into the sum of two quantities which depend on the bias and variance

c) What is meant by "unbiased" when referring to an estimator? Among all unbiased linear estimators for the regression problem for f, what is the remarkable property of β with respect to MSE?

d) Make an example of a biased estimator for linear regression which can be easily obtained from β by regularization, provide a formula for the estimator or mention an algorithm to obtain it, and comment on the advantages of employing such estimator in terms of MSE compared to β .