Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard
04/09/2012

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
- Either pen and pencil are allowed no other technical mean to support yourself is allowed

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 – 10 points)

With reference to the Linear Discriminant Analysis method/model for classification, describe:

1. Assumptions and analytical form of LDA
2. Its (dis)advantages wrt plain Linear Regression (on the indicator matrix), Logistic Regression, and Optimal Separating Hyperplanes
3. How do we train this classifier from data?
4. What is the difference between Linear Discriminant Analysis and Quadratic Discriminant Analysis
5. Does it work for multi-class problems? If yes how? If no, why?

Question 2: Decision Trees (Answer on sheet 1 – 6 points)

Describe the Decision Tree classification method; in particular:

1. The model used for classification
2. The algorithm (in details) to build such model from a data set
3. The issue of overfitting in Decision Trees and the ways to reduce it
4. Why should we turn a Decision Tree into a rule set?

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

Given the following algorithms:

1. k-means
2. Hierarchical
3. Mixture of Gaussians
4. DBSCAN
5. k-medoids

6. fuzzy c-means
7. Jarvis-Patrick

Complete the following sentences matching them with one (or more!) of the algorithms, answering the questions in parentheses and providing detailed explanations to motivate your choices.

(NOTE: although all the sentences refer to a single algorithm, there may be more than one valid choice. In these cases, provide and motivate all of them).

a) This algorithm relies on the concept of "neighbor" (what kind of problems is it better suited for?)

b) This algorithm is very fast, however it does not always find the best clustering (what would you do to make results better?)

c) This algorithm provides a "soft" classification (what does this mean?)

d) This algorithm can provide good results even if noise is present in the dataset (is it also able to detect which points are noise?)

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

Lets focus on the regression of a function y=f(X):

1. Describe the linear regression model for f, by defining both X and β (what are their dimensions?)
2. Define the Residual Sum of Squares. How can you minimize such quantity?
3. Derive the mathematical formula for the Least Squares Estimator β (beta_hat) and for y (y_hat). Write explicitly relevant passages.
4. Define the Mean Square Error of β (beta_hat) in estimating β, and derive a decomposition in terms of bias and variance.
5. Comment on the properties of bias and variance of the LSE. How does shrinkage and variable selection methods differ from LSE form this point of view?