Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard
15/07/2011

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 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 after the exam.

Question 1: Logistic Regression (Answer on sheet 1)

With reference to Logistic Regression describe:

1) Its assumptions and analytical form
2) its advantages wrt plain Linear Regression (on the indicator matrix), Linear Discriminant Analysis, and Optimal Separating Hyperplanes.
3) How do we train this classifier from data?
4) How does it work for multi-class problems?

Question 2: Decision Trees (Answer on sheet 1)

Describe the Decision Tree classification method; in particular:

1) The model used for classification
2) The algorithm (in details) to build such model from data
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: Regression (Answer on sheet 2)

1) Present and discuss the approach of shrinkage methods in linear regression.
2) Which algorithms belongs to this class? Give a brief presentation of each of them.
3) What are the advantages of such methods compared to other techniques?
4) Present the approach of ridge regression, write down both formulations for the minimization problem and the solutions, with the mathematical derivations.

5) Singular Values Decomposition gives some insights on the nature of ridge regression. Interpret ridge regression solutions from this point of view, present the main formula where SVD has been appied, and comment on the role of principal components and variance.

Explain the meaning of all letters you are using, and size of vectors and matrices.

Question 4: Clustering (Answer on sheet 3)

As you know, K-Means algorithm does not always give the best solutions you might expect. Anyway, it still has many applications.

1) Why?
2) Describe the limitations and the main advantages of K-Means
3) Explain which techniques you would use to make its results better.