Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard
19/09/2012

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: Logistic Regression (Answer on sheet 1 – 8 points)

With reference to the Logistic Regression method/model for classification, 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: Kernel Methods (Answer on sheet 1 – 8 points)

What is a Kernel Smoother and how it relates to k-nearest neighbors methods? What kernels are there? How can we use Kernel Smoothers for regression? And how for classification?

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

a) Explain in few sentences what cluster evaluation is, why it is needed, what is evaluated and why a "statistical framework" is often necessary to give an interpretation to evaluation results.

b) 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 preferable with respect to another one such as K-Means and when is it not?

c) Some algorithms (e.g., k-means, 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.

d) Describe in detail how the Fuzzy C-Means clustering algorithm works. In which parts is it similar and in which ones does it differ from K-Means?

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

1) Present the approach of variable subset selection in linear regression. How does it differ from ordinary least squares regression? When it may be useful in practice?

2) Introduce and discuss the three different algorithms that belong to this class of methods. What do they have in common and which are the main differences among them? Express explicitly for each algorithm, the condition according to which variables are dropped or added to the model

3) How can you determine the final number of variables to be included in the model in variable subset selection?

4) What are the main differences between subset selection methods and shrinkage methods?