Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard
10/09/2013

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 – 10 points)

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

a) Assumptions and analytical form of LDA

b) Its (dis)advantages wrt plain Linear Regression (on the indicator matrix), Logistic Regression, and Optimal Separating Hyperplanes

c) How do we train this classifier from data?

d) What is the difference between Linear Discriminant Analysis and Quadratic Discriminant Analysis

e) Does it work for multi-class problems? If yes how? If no, why?

Question 2: Kernel Smoothers (Answer on sheet 1 – 6 points)

Let's consider the use of local methods for Regression and Classification:

a) What is a Kernel Smoother and how it relates to the k-nearest neighbors method for regression and classification?

b) What kernels are there? Provide and comment two different ways to define the width of the kernel? How the size of the width could be decided?

c) How can we use Kernel Smoothers for regression? And how for classification? What is a Kernel density Estimator?

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

a) What is model-based clustering? Name a clustering algorithm belonging to this family and describe how it works.

b) What is the difference between a "fixed radius" neighborhood and a "self scaling" neighborhood? How are neighbors chosen? What are the advantages and disadvantages of these two different approaches?

c) Which algorithm would you use if you needed to find non-globular cluster in presence of noise? Name and describe the algorithm in detail.

d) Suppose you want to evaluate the clusters identified by an algorithm you created. You calculate entropy over the whole set of clusters and you obtain the value 0.7. What can you deduce from this?

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

a1) Present the detailed pseudo-code (for instance using matlab-style language) for the following two algorithms

-Forward stepwise

-Forward stagewise

(Notice that a general description of how the algorithms behave is not enough, you need to provide a complete description of the steps of each algorithm, i.e., you need to write the pseudo code implementation of the algorithm with X and y as input variables and theta as output variable. Just to be clear, do not simply write "choose the most correlated variable", but write the code or formula to identify such variable)

a2) What are the differences in terms of computational complexity of the two algorithms? which is the best suited for problems with large number of variables?

b1) Write down the closed form solution for the ridge regression estimator, and derive the formula (include all algebra!) starting from the minimization of the RSS according to the l2-penalized regression model or ridge regression

b2) What differs in lasso compared to ridge regression in the penalized formulation? Can you also provide a formula for the lasso estimator? If not, how can you obtain the lasso estimator for the regression problem?