Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard

20/09/2013

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: LDA and Logistic Regression (Answer on sheet 1 – 10 points)

Consider the Linear Discriminant Analysis (LDA) and Logistic Regression method/models for classification, describe:

a) Assumptions and analytical form of LDA and Logistic Regression

b) Should I use LDA or Logistic Regression for my classification task? Why?

c) How do we train these classifiers from data? Which one is the easiest/fastest to train ?

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

e) Do they 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?

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)

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

(8) Self-Organizing Maps

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 a "self-scaling" neighborhood (what does this mean? How can this be accomplished?)

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

(c) This algorithm builds new clusters by merging or splitting existing ones (describe the differences between the two approaches... And what about the complexity?)

(d) This algorithm works fine even with high-dimensional data (how does it reduce the dimensionality?)

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

1a) Present the detailed pseudo-code (for instance using matlab-style language) for the following two algorithms

-Best subset

-Backward stepwise selection

(Note 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 where X and y as the input variables and theta the output estimator. Just to be clear, do not simply write "choose the most correlated variable", but write the code or formula to identify such variable)

1b) What are the differences in terms of computational complexity of the two algorithms above? which is the better suited for problems with large number of variables?

2a) 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

2b) 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?