Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard

26/07/2013

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

a) Describe Logistic Regression (LR), its analytical form, its assumptions, and its advantages, if any, wrt Linear Discriminant Analysis.

b) Derive the LR (linear) decision boundary among classes from the Logistic Regression analytical formula

c) Describe the training algorithm for LR in the case of two classes.

d) Is Logistic Regression suitable for multi class classification problems or the linear boundary implies the use of two classes? How can we deal with multiple classes problems?

Question 2: Kernel Smoothers (Answer on sheet 1 – 8 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)

Suppose you want to evaluate the results of some clustering algorithms using SSE and Accuracy.

1) Which of these measure is defined as "internal", which is "external", and what does this mean?

2) You plan to develop your own functions to compute SSE and accuracy. What would be the input and the output of each of these functions?

3) One of the clustering algorithms allows you to choose the number of clusters in advance. You calculate SSE after different executions of this algorithm, using

K=2,3,...,10. SSE for K=10 provides the lowest value: what can you deduce from this?

4) Now suppose you have ground truth for your dataset. You run two different clustering algorithms on the same data and obtain the following results:

	SSE	Accuracy
Algorithm1	115.3	87%
Algorithm2	1285	95%

What is the meaning of these results? Which algorithm is better?

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

1) What is the purpose of variable selection in linear regression? Give at least two different motivations, one related to the mean square error decomposition and the other related to the point of view of the final user who is going to use the regression model to prediction the value of y, given x.

2) How variable selection can be achieved in linear regression? Mention two different approaches to variable selection, and describe with sufficient details one algorithm for each class. (Hint: for each of the approaches I refer to, we have seen more than one algorithm. In other words with "approach" I don't refer to a specific algorithm, instead I refer to what the different algorithms in that the class have in common)

3) How can you chose the number of variables when you apply variable selection techniques? Can you choose that a priori? Refer explicitly to how the number of variables in the regression model is controlled for the two algorithms you have presented in 2)

4) What variable selection and shrinkage methods have in common? Why these techniques sometimes may be more useful than ordinary least squares?