POLITECNICO DI MILANO
Dipartimento di Elettronica e Informazione

Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard
03/02/2012

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 write it on top of your assignment and ***tell it to the teacher!***.

Question 1: Linear Classification (Answer on sheet 1)

Describe shortly Linear Regression and how we can use plain Linear Regression for classification. What is its computational complexity? What to we mean by "hiding effect" of Linear Regression on the Indicator Matrix? How can we avoid it?

Question 2: Local Regression Methods (Answer on sheet 1)

Describe the idea behind Local Linear Regression and Local Polinomial Regression. What do we mean by "Equivalent Kernel" in Local Regression? How can we select the with of the kernel used for Local Regression?

Question 3: Clustering (Answer on sheet 2)

K-means is a clustering algorithm that, despite some limitations, is still widely used for many applications.

1) Highlight the main advantages of using K-Means instead of another clustering algorithm (you can explicitly compare K-Means with other algorithms you choose) and suggest some applications in which you consider it better.

2) What approach would you suggest to address the fact that the result of K-Means clustering depends on the initial positions of centroids?

3) What approach would you suggest to address the need of knowing the number of clusters in advance?

4) What clustering algorithm would you suggest to address K-means limit of not being able to deal with non-globular clusters? Choose one (if there are many) and motivate your answer with respect to K-Means.

Question 4: Regression (Answer on sheet 3)

1) Explain differences and similarities between ridge regression and the lasso. Which minimization problems do you need to solve in each method?
2) Present and discuss the Least Angle Regression pseudo-code