Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard
29/01/2013

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
- 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)

Describe Linear Discriminant Analysis (LDA), its analytical form, its assumptions, and its advantages wrt plain Linear Regression (on the indicator matrix). How does it work for multi-class problems? Describe the two techniques to implement Quadratic Discriminant Analysis?

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)

Given the following algorithms:

(1) k-means

(2) Hierarchical

(3) Spectral Clustering

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

(c) This algorithm can provide good results even when non-globular clusters are present (how does it/do they work?)

(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)

Discuss differences and similarities between ridge regression and the lasso:

1) Which optimization problem do you need to solve in each method? Write down both the constrained formulation and the formulation with the penalizing term. Give a geometrical interpretation for the constrained formulation.

2) Can you provide a closed-form solution to any of the optimization problems in the previous point? If yes, derive β mathematically, starting from X matrix (N observations and m variables) and y vector (N observations)

3) If the X variables are orthonormal, there is an explicit solution for β of lasso and β of ridge regression, expressed as a simple transformations of the $\tilde{\beta}$ obtained with least squares. Explain the two different effects, and represent them on a graph where on the x axis you have $\tilde{\beta}$ of ordinary least squares, and on the y axis β of lasso and β of ridge regression.

4) Describe a real world scenario in which lasso would be more suitable then ridge regression and don't forget to justify you answer.