POLITECNICO DI MILANO
Dipartimento di Elettronica e Informazione

Pattern Analysis and Machine Intelligence

M. Matteucci, L. Malagò, D. Eynard
19/09/2011

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 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 the Decision Tree classification method; in particular:

1. The model used for classification
2. The algorithm (in details) to build such model from data
3. The issue of overfitting in Decision Trees and the ways to reduce it
4. Why should we turn a Decision Tree into a rule set? How can we do it?
5. How does the chi squared independency test work? How can it be used to post-prune a decision tree?

Question 2: Cross-Validation and Overfitting (Answer on sheet 1)

What do we mean by model assessment? And what aboutmodel selection? What is the Bias-Variance decomposition and how it is related to model complexity? What are cross-validation and k-fold cross-validation?

Question 3: Clustering (Answer on sheet 2)

Hierarchical clustering is not a single algorithm but rather a family of different clustering algorithms. Explain:

1. how this family is composed
2. how these algorithms work
3. what different metrics exist to measure the distance between clusters
4. when is a hierarchical algorithm preferrable with respect to another one such as K-Means and when is it not?

Question 4: Regression (Answer on sheet 3)

Let $X^T = (X_1, X_2, ..., X_p)$ an input vector and Y a real-valued vector

1. Write the form of the linear regression model (0.5 points over 8)

Let $(x_1, y_1), ..., (x_N, y_N)$ a training set

2. write the least squares formula, i.e., the residual sum of squares as a function of \hat{β} (0.5 points over 8)
3. from the formula of residual sum of squares derive mathematically the solution for \hat{β}. Use matrix notation, where X is the Nx(p+1) matrix of the input vectors and y the N-dimensional vector of outputs. (2 points over 8)
4. write the formula of $\hat{y}$, the prediction of y given by the the linear regression model (0.5 points over 8)

Consider now the univariate case, when $p=1$, with no intercept

5. write the formula for $\hat{\beta_1}$ and the residuals (1 points over 8)

Consider now the univariate case, when p=1, with an intercept

6. write the formula for $\hat{\beta_1}$ and interpret it in terms of orthogonalization and regression (1 points over 8)

Suppose next that the the inputs $x_1, x_2, ..., x_p$ are orthogonal

7. write formula for $\hat{\beta_j}$, with $1 \leq j \leq p$ (0.5 points over 8)
8. describe the pseudo code of the regression by successive orthogonalization (2 points over 8)