Assignment 5: Performance Evaluation

Submission: Thursday June 18th
Maximum of 2 students per group

Prof. Fabio A. González
Machine Learning - 2009-I
Maestría en Ing. de Sistemas y Computación

1. [Alp04] Exercise 2 (chap. 14, page 349)

2. Evaluating the performance of a supervised learning algorithm:

   (a) Download the Wisconsin Breast Cancer data set from `http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)` and divide it in a training set and a test set (70/30).

   (b) Use a decision-tree learning algorithm to learn a model.

   (c) Find a confidence interval for the expected error of the trained model.

   (d) Use 10-fold cross validation to evaluate the performance of a decision tree and a SVM.

   (e) Use a K-fold cross-validated paired t-test to determine whether the two models have the same error rate or not.

3. [TSK05] Exercise 23 (chap. 8, page 565)

4. [TSK05] Exercise 24 (chap. 8, page 565)

5. [TSK05] Exercise 27 (chap. 8, page 566)

References

[Alp04]   Alpaydin, E. 2004 Introduction to Machine Learning (Adaptive Computation and Machine Learning). The MIT Press.

[TSK05] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, 2005, Introduction to Data Mining, Addison-Wesley.