CS909: 2013-14

Week 8: Classifier evaluation

1. Suppose that you have used some concept learning algorithm to learn a hypothesis h1 from some training data. You are interested in knowing the accuracy that the hypothesis can be expected to achieve on the underlying population. You assess the hypothesis on a set of test data consisting of 45 instances and you observe an error of 6.67%. Calculate the 95% interval for the expected error.

2. You compare two supervised learning classification algorithms on a set of training data and you find out that 10-fold cross validation yields the following accuracies:

CV Fold	Algorithm 1	Algorithm 2
1	91.11	90.7
2	90.48	90.52
3	91.87	90.88
4	90.52	90.87
5	89.88	90.02
6	89.77	88.99
7	91.44	90.98
8	90.88	91.44
9	90.77	90.77
10	90.89	90.92

At what confidence level can you assume that algorithm 1 will outperform algorithm 2? Show your working.

3. Ravensworth, Liakata and Clare [1] trained a classifier on predicting the type of a scientific article (e.g. Review, Research etc.) according to PlosOne categories and using as features the distribution of Core Scientific concepts within papers (e.g. Hypothesis, Methods, Conclusion etc.).

    Perform the same classification task on the data set AI2013_papers.csv using random forests, naiveBayes and SVM from the e1071 R package.

    Compare classifier performance using 10-fold cross-validation and report appropriate measures. Give confidence levels for the accuracy of the best performing classifier.(If you get a higher F-score than the paper on some of the classes please let us know!)

    [1] J. Ravenscroft, M. Liakata, A. Clare (2013). Partridge: An effective system for the automatic classification of the types of academic papers. In Proceedings of AI-2013.

Submission deadline: Midday, Thursday 6th March 2014