

Supervised classification methods

Pedro Larrañaga¹

¹ *Universidad Politécnica de Madrid, pedro.larranaga@fi.upm.es*

The tutorial will cover some of the most popular supervised classification methods: non-probabilistic, such as K-nearest neighbours and classification trees; probabilistic, such as logistic regression and Bayesian classifiers and also metaclassifiers. For each method the general algorithm as well as some of its variants will be explained.

Honest validation methods --repeated training and test, k-fold cross-validation and bootstrap-- for estimating the performance of a classifier will be discussed. Guidelines about their use will be presented. Several performance metrics as the accuracy, sensitivity, specificity, F1 measure, and the area under the ROC curve will be shown.

Feature subset selection based on filter (univariate and multivariate) and wrapper approaches for selecting relevant and non-redundant variables will be discussed.

The attendees of the tutorial will do some practical exercises with WEKA software applied to a real microarray data set.