Descriptions of classification problems

Classification problem 1: Fisher

The dataset of the first problem gives “Ronald Fisher's measurements of *type*, *petal width* (PW), *petal length* (PL), *sepal width* (SW), and *sepal length* (SL) for a sample of 150 irises. The lengths are measured in millimeters. Type 0 is *Setosa*; type 1 is *Verginica*; and type 2 is *Versicolor”*.

This problem is interesting because it describes a classic pattern recognize problem in nature and can be deployed to tons of similar classification problems. In addition, the dataset has the feature that “One class is linearly separable from the other 2; the latter are NOT linearly separable from each other”. This classic dataset has been adopted to compare the performances of different classifiers.

The data set is available from UCI’s Machine Learning Repository online:

<http://archive.ics.uci.edu/ml/datasets/Iris>

The data set has already been prepared for learning, uploaded as a matrix form in Matlab.

Classification problem 2: Wine

“These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. In a classification context, this is a well posed problem with "well behaved" class structures.” Different from problem 1, all attributes are continuous in this case.

The data set is available from UCI’s Machine Learning Repository online:

<http://archive.ics.uci.edu/ml/datasets/wine>

The data set has already been prepared for learning, uploaded as a matrix form in Matlab. The first attribute is the class from 1-3.

Two classification problems were tested by five methods: Decision Trees; Neural Networks; Boosted Decision Trees; Support Vector Machines (SVM); K-Nearest Neighbors. The results were compared from various aspects.

Decision Trees

Neural Networks

Boosting

Support Vector Machines

K-Nearest Neighbors