

Machine Learning

Exercise 5

Marc Toussaint

Machine Learning & Robotics lab, U Stuttgart
Universitätsstrae 38, 70569 Stuttgart, Germany

June 3, 2013

1 WEKA, decision trees and boosting

Weka <http://www.cs.waikato.ac.nz/ml/weka/> is a Java toolbox for Machine Learning. Download and install it (an Ubuntu package exists).

Your task is to test various classifiers on the MNIST digit data set from <http://yann.lecun.com/exdb/mnist/>.

a) Download the datasets `digits-train.arff.gz` and `digits-test.arff.gz` from the course webpage. This is the result of a (painful) conversion of the original datasets to the arff format. `gunzip` the data. Start Weka with `weka -m 2g` (or more memory, otherwise you'll not be able to load the dataset. In explorer mode, load the dataset.

b) Try to find the best possible classifier using Weka. Focus on ensemble and decision tree methods. (You can choose from a large range of classifiers and change their parameters by right click on their name...) Compare with the results reported at <http://yann.lecun.com/exdb/mnist/>

c) Weka also allows you several preprocessing filters, including Principle Components. Test whether you can get better results using this preprocessing.