Machine Learning 2015: Project 2 - Classification Report

trubeli@student.ethz.ch
tdenoreaz@student.ethz.ch
liliu@student.ethz.ch

November 19, 2015

Experimental Protocol

We started by using a multi-class svm on the raw data. We first wrote and experiment with the sci-kit learn classifier classes in order to obtain prediction. In a second step, we tried to enhance our score by normalizing the training dataset by:

1. Subtracting the mean of each feature of the dataset
2. Divide each feature vector by its standard deviation.

This normalization however seemed to fail in our case and lead to worst results. Finally we also tried to smooth the data points using a Gaussian filter. The idea behind this was to smooth the data and removing the possible outliers values.

1 Tools

We did most of the processing using Matlab, in the early stage of the project we used the sci-kit learn package in Python.

2 Algorithm

We used two algorithms during our tests:

2.1 Support Vector Machine

We set the kernel using the *Gaussian radial basis function*. However the performance were not as better as what we have using Random Forest.

2.2 Random Forest

The Random Forest Classification Algorithm. However, in order to process the features before feeding it to the algorithm. We decided to apply a Gaussian Filter in order to smooth the data and hopefully reduce the error.

3 Parameters

We used 20 trees for our Random Forest algorithm.

4 Lessons Learned

The treebagger class of matlab seems to perform better compared to the random forest package in scikit learn. We have also noticed that the use of random forest seemed to always yield better parameters for our model.