Introduction to Learning and Intelligent Systems – Spring 2015

taubnert@student.ethz.ch
junkerp@student.ethz.ch
kellersu@student.ethz.ch

March 15, 2015

1 Project Regression – Team "awesome"

1.1 logscore

Since we can not use logscore directly as a distance function during regression, we need to transform our data x, y. Suppose we want to search a function $f(x)$. Then to minimize $logscore(f(x), y)$ we minimize the two-norm $||f'(x) - y'||_2$ instead. Looking at the definition of $logscore$, we see that this can be accomplished by choosing $f'(x) = \log(1 + f(x))$ and $y' = \log(1 + y)$. The function f can then be reconstructed by $f(x) = \exp(f'(x)) - 1$.

1.2 Regressors

We used a number of different regressors. Most of them we understand how they work but we also used the *RandomForestRegressor* which we don't understand at all.

In the end, we compared a simple linear regression, a ridge regression, a k nearest neighbors regression, a lasso regression with the random forest regression. We concluded that we can do almost as good as the random forest regression.

For some parameters of the regressors, e. g. the α value for ridge regression and the k value for k nearest neighbors, we used a simple grid search.

1.3 Features

Different heuristics lead us to use different basis functions for our features. Because the data is about train usage and we have a timestamp provided, we assumed that there will be some periodicity observeable. This assumption let us add the fourier and also the discrete cosine transformation as base-functions.

We deliberately didn't use all of the available data to fit. We concluded that time in the minute frequency and the weather data D corresponds mostly to noise.

1.4 Scores

The scores with our final run for generating the test set are the following. Because we scored best with the random forest generator, we decided to go with that one, although it seem to have overfit to our training data.

Linear	0.4249	0.4326	0.4460
Ridge	0.4280	0.4354	0.4477
Random Forest	0.1619	0.1606	0.3925
K-NN	0.4949	0.6278	0.6047
Lasso	0.4276	0.4345	0.4477