

# Introduction to Learning and Intelligent Systems - Spring 2015

taubnert@student.ethz.ch  
philj@student.ethz.ch  
susik@student.ethz.ch

March 15, 2015

## 1 Project Regression

### 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  $\text{logscore}(f(x), y)$  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-neighbours regression, a lasso regression with the random forest regression. We concluded that we can do almost as good as the random forest regression.

### 1.3 Features

Different heuristics lead us to use different basis functions for our features.