

Machine Learning 2014: Project 1 - Regression Report

trubeli@student.ethz.ch
tdenoreaz@student.ethz.ch
marmarti@student.ethz.ch

October 31, 2015

Experimental Protocol

We started by plotting the correlation matrix in Figure 1. This was used in order to have a grasp of how the features were correlated to each other.

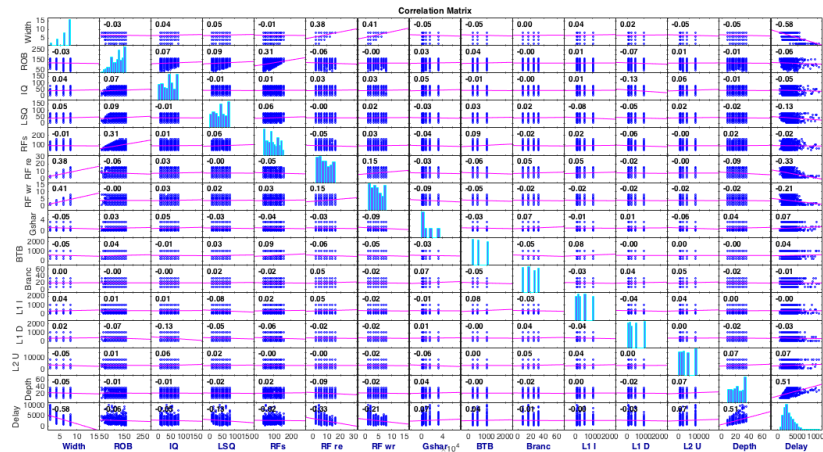


Figure 1: Correlation Matrix

1 Tools

We did most of the processing using Matlab we also use R with the package *randomForest* 4.6-12.

2 Algorithm

Linear Regression

We started by implementing a first and simple version of the linear regression to have a starting point. Using the correlation matrix in Figure 1, we picked some of the features to establish our model.

K-Fold test

In order to try different algorithm and to be sure they improved. We implemented a *K-Fold test* which compute the *RMSE factor* against each *k*-part of the training set and then plotted it to have the average *RMSE* and also the variance.

LASSO

We used the Matlab *lasso* function in order to compute the regression based on LASSO. To do so, we used also 10 cross validations and we limited the parameter $\lambda \in [3, 5]$.

Random Forest

We decided to implement the Random Forest algorithm in *Matlab* and also in *R*. This was mainly done by

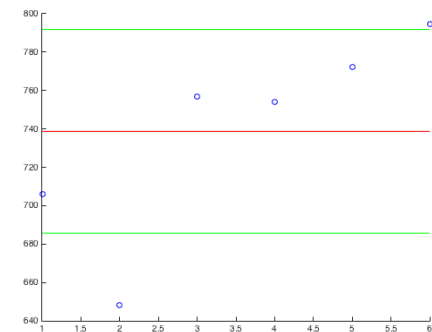


Figure 2: K-Fold test

3 Features

We always started by using the linear model, and then added some of the quadratic terms for each of the algorithms.

4 Parameters

We used 10-fold cross validation to search for λ in LASSO.

5 Lessons Learned

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