Machine Learning 2014: Project 1 - Regression Report

vlucas@student.ethz.ch
ivankaya@student.ethz.ch
piusv@student.ethz.ch

October 31, 2015

Experimental Protocol

1 Tools

We exclusively used $Matlab$ for this assignment. We used $csvread$ to read the provided training and validation data sets and $csvwrite$ to write our predictions. For regression and crossvalidation we used the $lasso$ function. The $plot$ function was also useful for data exploration and guessing useful non-linear feature transformations. The figure in section 4 was created with $lassoPlot$.

2 Algorithm

We used $L1$ regularized least squares regression provided by the $lasso$ function in $Matlab$.

3 Features

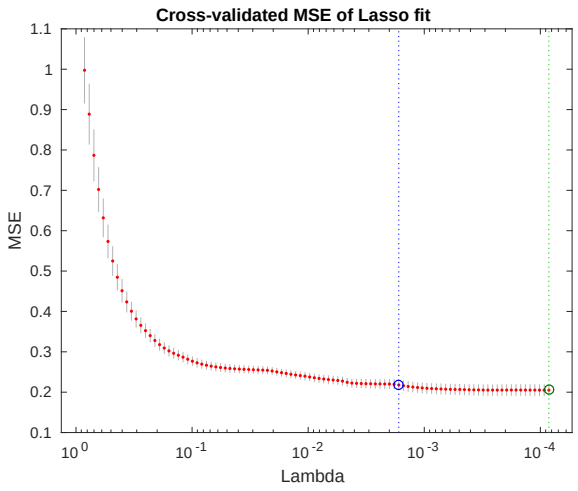
From the original set of 14 features $\{x_1, x_2, \ldots, x_{14}\}$ we constructed the new features

$$\left\{z_1, \ldots z_5, z_1^2, z_1 z_2, \ldots, z_4 z_5, z_5^2, z_1^3, z_2^3, \ldots, z_5^3\right\}$$

where $z_1 = \frac{1}{x_1}$, $z_2 = x_2$, $z_3 = x_4$, $z_4 = \frac{1}{x_6}$, $z_5 = x_{14}$.

4 Parameters

We used lasso regression with 10 fold cross validation to determine our model parameters. The $lasso$ function provided by $Matlab$ allows us to do so easily. Additionally it lets us find the $L1$ penalty parameter λ which gives the lowest expected cross validation error.



5 Lessons Learned

Additionally to lasso regression we tried training a Neural Network (provided by a $Matlab$ toolbox) and Gaussian Process regression (using the open source $GPML$ library). They both seemed to perform worse because of a lack of additional training data.