

# Machine Learning 2015: Project 1 - Regression Report

pungast@student.ethz.ch

October 27, 2015

## Experimental Protocol

To obtain the model from raw feature vector  $X = (x_1, \dots, x_n)$ , the following steps were taken. Each step was applied to the result of last step.

1. A subset of the original features were removed. Selection criteria were similar to those laid out in Step 5.
2. Nonlinear combinations of the original features were added (described in Section 3).
3. Features were normalised by subtracting the mean and dividing by the standard deviation.
4. A bias term was added.
5. Redundant features were identified by removing each feature  $x_i$ , running 10-fold cross-validation on features  $X' = X \setminus x_i$ . A feature was declared redundant if the mean RMSE from cross-validation on features  $X'$  decreased by 0.05 compared to using all features.
6. A **ridge regression** model with normalisation parameter  $\lambda = 0.1$  was trained. The value of  $\lambda$  was selected in 10-fold cross-validation to produce the lowest mean RMSE.
7. Weights  $(w_1, \dots, w_n)$  of the model were used to find the predicted value  $y = \sum_{i=1}^n w_i x_i$ .

## 1 Tools

I used:

- Python 2.7
- scikit-learn for training a LASSO classifier
- NumPy for everything else

All code used for this project is wrapped into a single class in `Regressor.py`.

## 2 Algorithm

Ridge regression was chosen as the best model in this project. Describe the algorithm you used for regression (e.g. ordinary least squares, ridge regression, ...)

## 3 Features

The following features were added:

- a bias term  $x_0 = 1$
- $\log(1 + x_i)$
- $x_i \log(x_i)$
- $\sqrt{x_i}$
- $x_i x_j$  for all  $i, j \in \{1, \dots, n\}$
- $x_i^3$
- $x_i^4$

## 4 Parameters

How did you find the parameters of your model? (What parameters have you searched over, cross validation procedure, ...)

## 5 Lessons Learned

What other algorithms, tools or methods did you try out that didn't work well? Why do you think they performed worse than what you used for your final submission?