

Machine Learning 2015: Project 1 - Regression Report

pungast@student.ethz.ch

October 28, 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 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 x_0 was added to the data.
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 or more compared to using all features.
6. A **ridge regression** model with normalisation parameter $\lambda = 0.1$ was trained. The value of λ was selected in 10-fold cross-validation to produce the lowest mean RMSE.
7. Weights (w_0, \dots, w_n) of the model were used to find the predicted value $y = \sum_{i=0}^n w_i x_i$.

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.

3 Features

The following features were added:

- 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

To find the hyperparameter λ of ridge regression, I used 10-fold cross-validation, testing values of lambda ranging from 0.01 to 10000, with each tested lambda value roughly 3 times larger than the previous one ($[0.01, 0.03, 0.1, \dots]$). The goal was the lowest mean RMSE over all 10 folds.

5 Lessons Learned

In addition to ridge regression, I tested LASSO regression. The CV error for LASSO was higher than for ridge regression. However, I did not conduct a formal parameter sweep (for LASSO) since ridge regression already had a quite low RMSE of 670. It is entirely possible that LASSO would perform better given the right regularisation parameter.

At the start of the project, it was clear that adding nonlinear features decreased the RMSE on the training set. However, with too many features the model overfit (CV error was high while training error was low). Conclusion: it is very important to remove redundant features to avoid overfitting the model.