# Machine Learning Assignment 3

Ryan Cooper

May 5, 2017

The first part of the assignment asked for linear regression and ridge regression to be done using the closed form solution where

$$w = (\|X\|_2^2 + \lambda I)^{-1} X^T y$$

$X$ represents the data to regress on and $y$ represents the data labels. $w$ is the weight vector, and $I$ is the identity matrix. $\lambda$ is the hyperparameter that represents the gain on the regularization term in the loss function

$$L = \|y - Xw\|_2^2 + \lambda \|w\|_2^2$$

$\lambda = 0$ for linear regression. For ridge regression $\lambda$ was chosen by finding the $\lambda$ that gave the smallest validation error rate, and selecting $\lambda$s based on the following guidelines.

> "You can begin by running the solver with $\lambda = 400$. Then, cut $\lambda$ down by a factor of 2 and run again. Continue the process of cutting $\lambda$ by a factor of 2 until you have models for 10 values of $\lambda$ in total."

Gradient descent was done with a learning rate of $10^{-6}$ which was chosen by a process of trial and error. Convergence was defined as

> "the change in any coefficient between one iteration and the next is no larger than $10^{-5}$."

Quotes were taken from the assignment description.

The output of the program is given below

> Linear Regression:
> Training RMSE = 0.797519498089
> Testing RMSE = 0.835386861685
>
> Ridge Regression:
> Training RMSE = 0.803719050184
> Testing RMSE = 0.84216331431
>
> Linear Regression with Gradient Descent:
> Training RMSE = 0.819324569049

Testing RMSE = 0.849747523083

Ridge Regression with Gradient Descent:
Training RMSE = 0.818521108773
Testing RMSE = 0.864392768628

Github repo: `https://github.com/Pillager225/Regression.git`