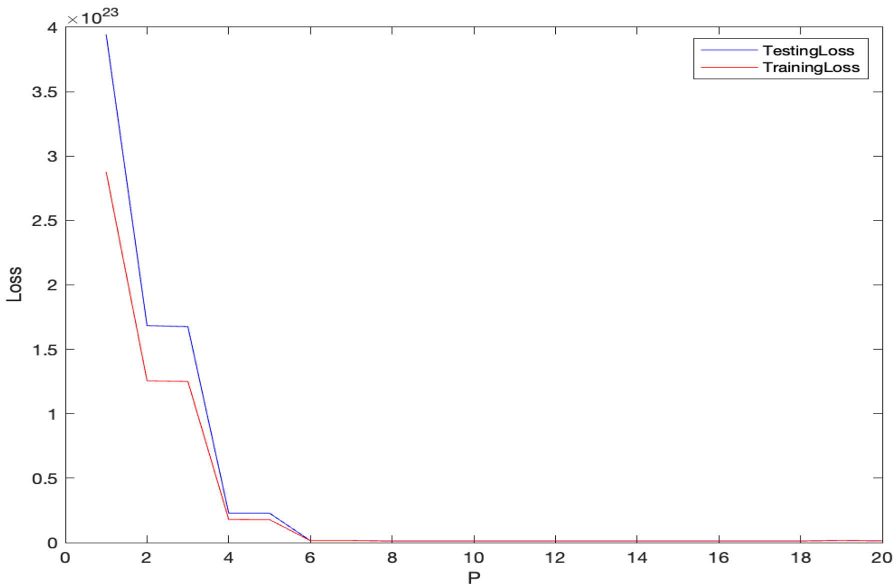
ML Assignment 1

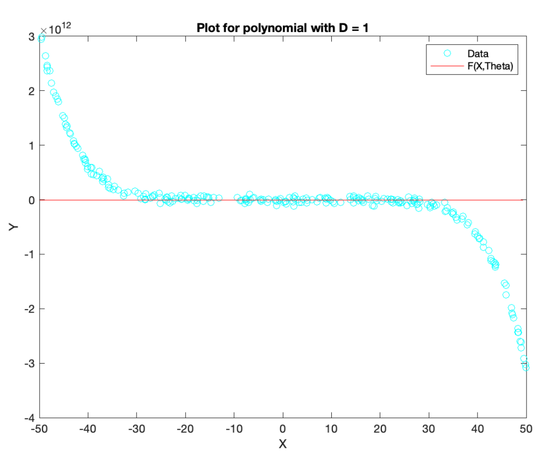
Simran Kucheria (sk11645@nyu.edu)

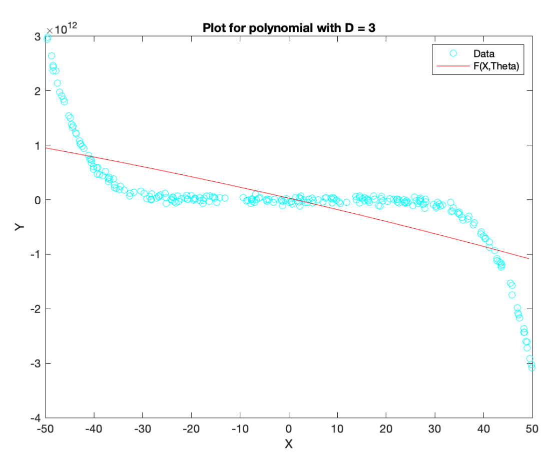
Problem 1

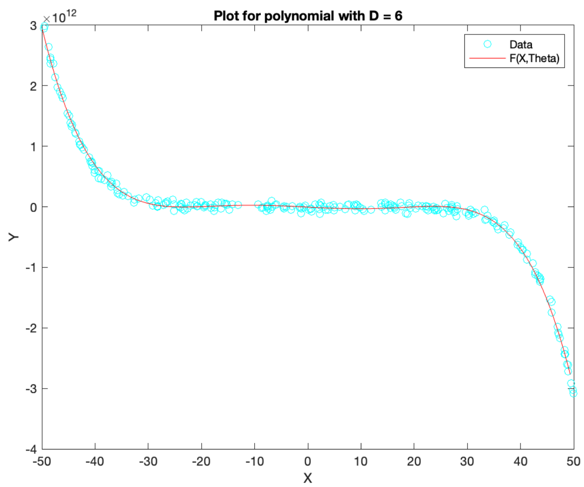
Used cross validation to split the data into 2 parts and calculated loss over various values of the d.
Plotting a graph for Loss with respect to the values of the polynomial d.

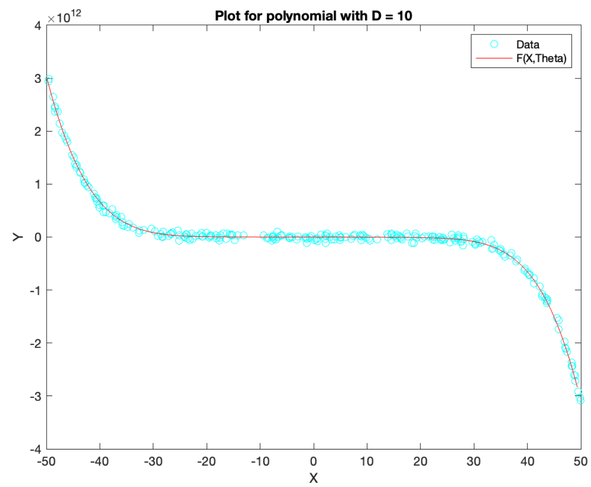


The Testing Loss reaches a minimum and starts stabilizing after **d = 6** making it the optimum value.
Plotting the polynomial function over the data for values of d = [1,3,6,10]





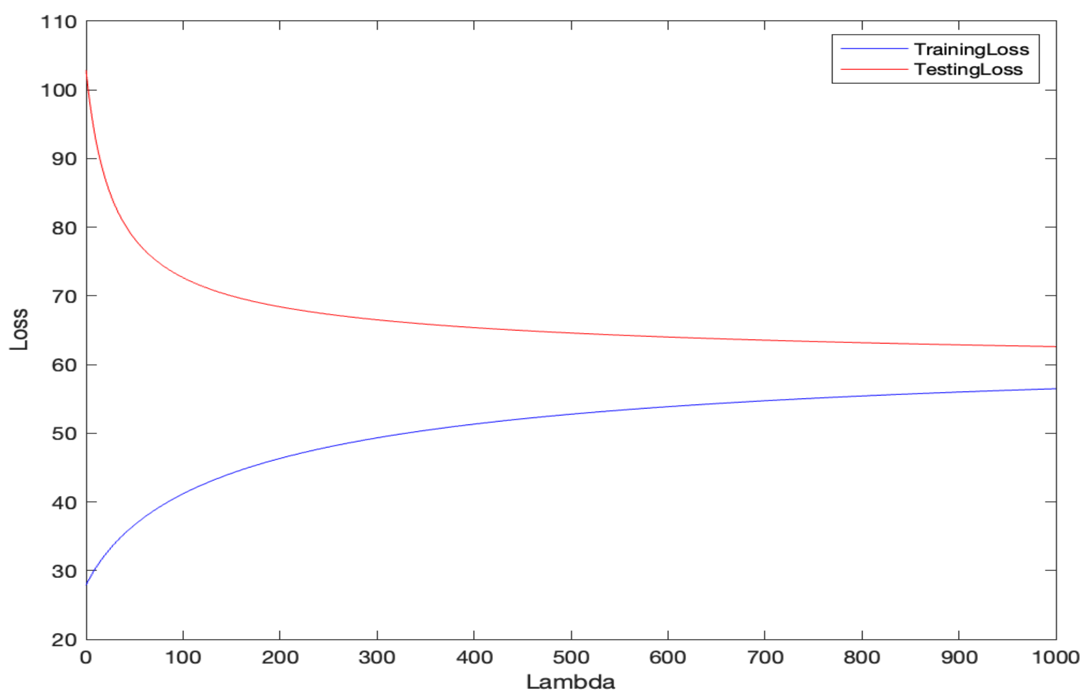




Problem 2

Regularised the risk with l2 regularisation using lambda. Split the data set into 2 equal halves to plot loss over testing and training data with respect to lambda.

Plotting a graph for the training and testing loss across different values of Lambda



We can approximate the value of **Lambda ~ 700** and the Testing loss minimizes with negligible difference at Lambda >= 700

Problem 3

Q3).

$$g(z) = \frac{1}{1+e^{-z}}$$

Prove that

$$g(-z) = 1 - g(z).$$

$$g(-z) = \frac{1}{1+e^{z}}$$

$$1 - g(z) = 1 - \frac{1}{1+e^{-z}}$$

$$= \frac{1+e^{-z}-1}{1+e^{-z}}$$

$$= \frac{e^{-z}}{1+e^{-z}} \cdot \frac{\times e^{z}}{\times e^{z}}$$

$$= \frac{1}{e^{z}+1} = g(-z).$$

Problem 4

Deriving the gradient of the risk function

Q4

$$R_{emp}\,\theta = \frac{1}{N}\sum_{i=1}^{N} (y_i - 1)\log(1 - f(x_i;\theta)) - y_i(\log(f(x_i;\theta)))$$

$$f(x,\theta) = \frac{1}{1+e^{-\theta^T x}}$$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\frac{d\sigma(x)}{dx} = \left(\frac{1}{1+e^{-x}}\right)' = -\frac{(1+e^{-x})'}{(1+e^{-x})^2} = -\frac{-e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1+e^{-x}-1}{(1+e^{-x})^2} = \frac{1}{(1+e^{-x})} - \frac{1}{(1+e^{-x})^2} = \frac{1}{(1+e^{-x})}\left(1 - \frac{1}{1+e^{-x}}\right)$$

$$= \sigma(x)(1-\sigma(x))$$

$$\frac{dR}{d\theta} = \frac{-1}{N}\sum_{i=1}^{N}\left[y_{(i)} * \frac{1}{f(x,\theta)} * \frac{d(f(x,\theta))}{d\theta}\right] + \sum_{i=1}^{N}(1-y_i) * \frac{1}{1-f(x,\theta)} * \frac{d(1-f(x,\theta))}{d\theta}$$

$$= \frac{-1}{N}\sum_{i=1}^{N} y_i * \frac{1}{f(x,\theta)} * \sigma(x)(1-\sigma(x)) * \frac{d\theta^T x}{d\theta} + \sum_{i=1}^{N}(1-y_i) * \frac{1}{(1-f(x,\theta))}$$

$$\hookrightarrow * -\sigma(x)(1-\sigma x)\cdot\frac{d\theta^T x}{d\theta}$$

$$\frac{-1}{N}\sum_{i=1}^{N}\left[y_{(i)} * \frac{1}{\cancel{f(x,\theta)}} * \cancel{f(x,\theta)}(1-f(x,\theta)) * \frac{d\theta^T x}{d\theta} + \sum_{i=1}^{N}(1-y_i) * \frac{1}{\cancel{1-f(x,\theta)}} * -f(x,\theta) * \cancel{(1-f(x,\theta))} * \frac{d\theta^T x}{d\theta}\right.$$

$$= \frac{-1}{N}\sum_{i=1}^{N} y_{(i)} * (1-f(x,\theta)) * \frac{d\theta^T x}{d\theta} + \sum_{i=1}^{N}(1-y_i) * -f(x,\theta) * \frac{d\theta^T x}{d\theta}$$

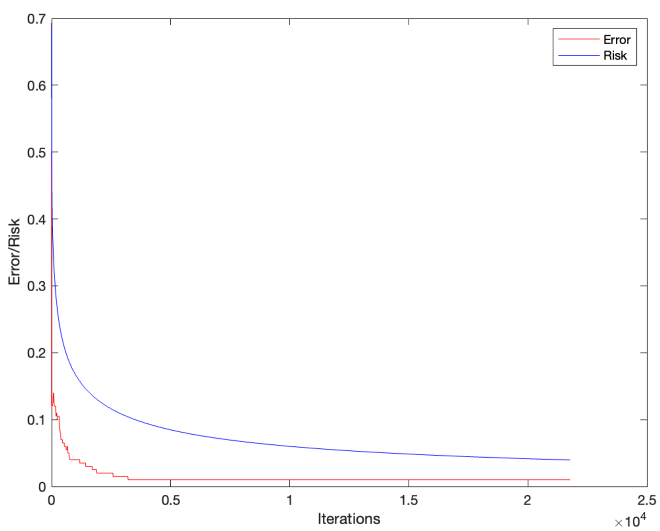
$$= \frac{-1}{N}\sum_{i=1}^{N}\left[y_{(i)} - y_{(i)} f(x,\theta) + y_{(i)} f(x,\theta) - f(x,\theta) \right] \frac{d\theta^T x}{d\theta}$$

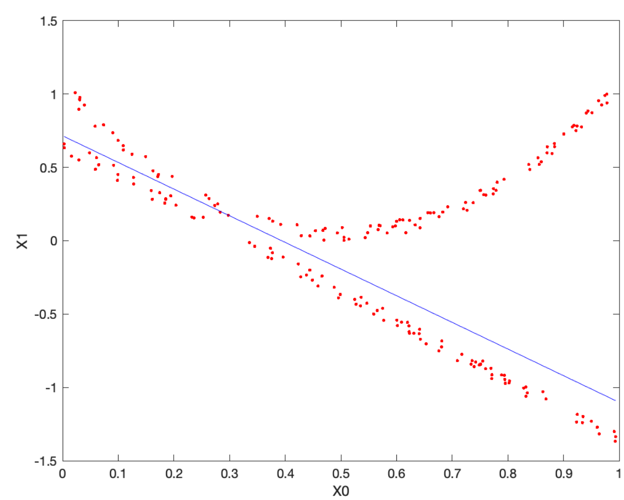
$$= \frac{-1}{N}\sum_{i=1}^{N}\left[y_{(i)} - f(x_i,\theta) \right] x^T$$

$$= \frac{-1}{N}\left(Y - f(X) \right) X^T$$

Taking different values of l and epsilon

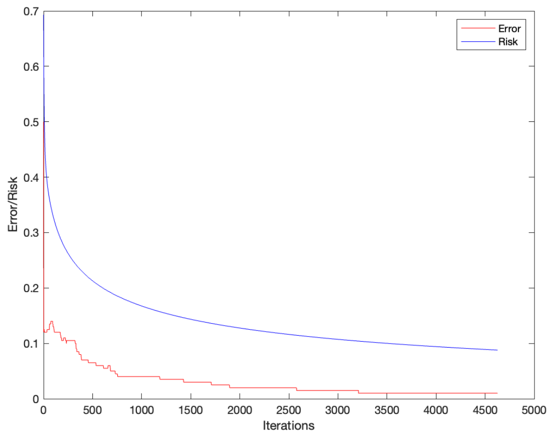
L = 1, Epsilon = 0.001

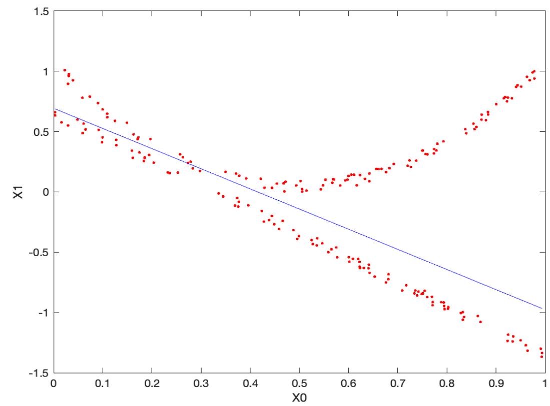




Gradient Descent converges at the **21778th iteration**

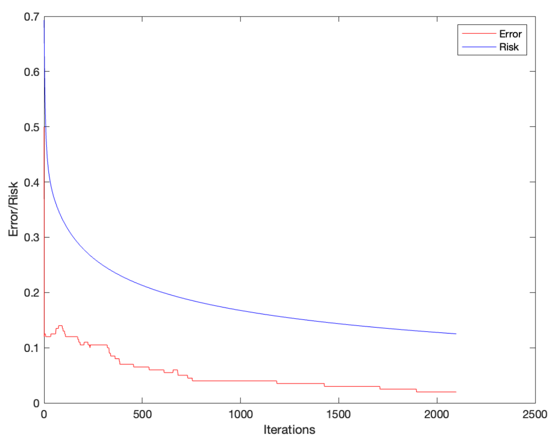
L = 1, Epsilon = 0.003

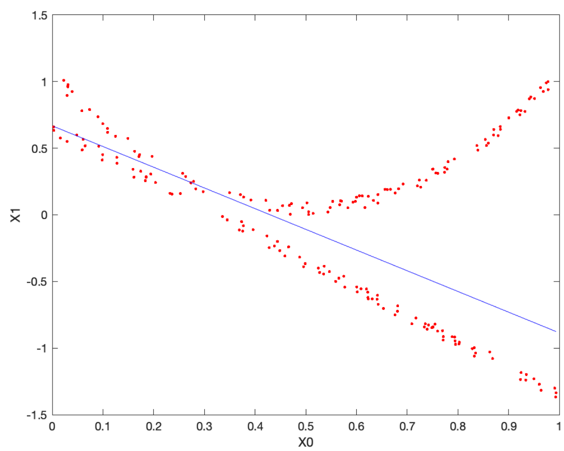




Gradient Descent converges at the **4626th iteration**

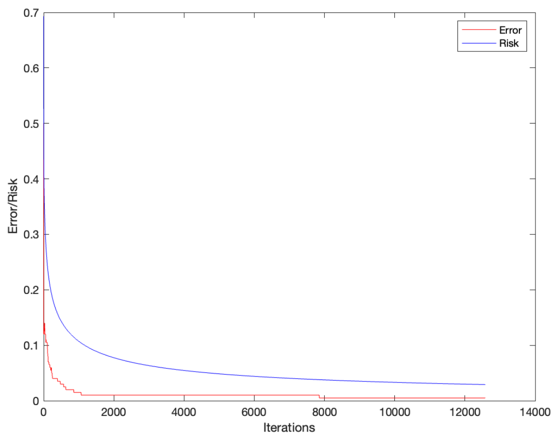
L = 1, Epsilon = 0.005

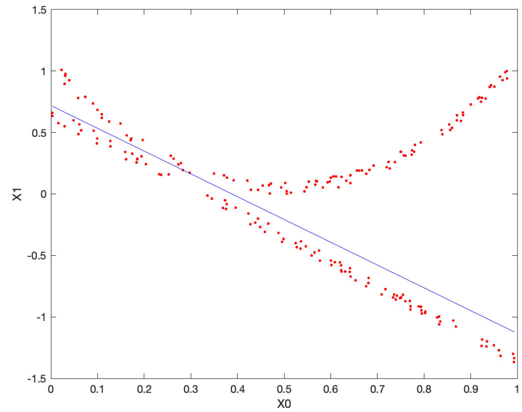




Gradient Descent converges at **2098th iteration**

L=3, Epsilon = 0.002





Gradient Descent converges at **12579th iteration**