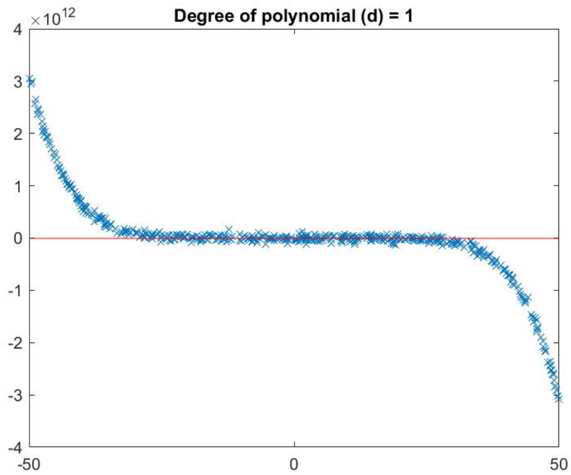
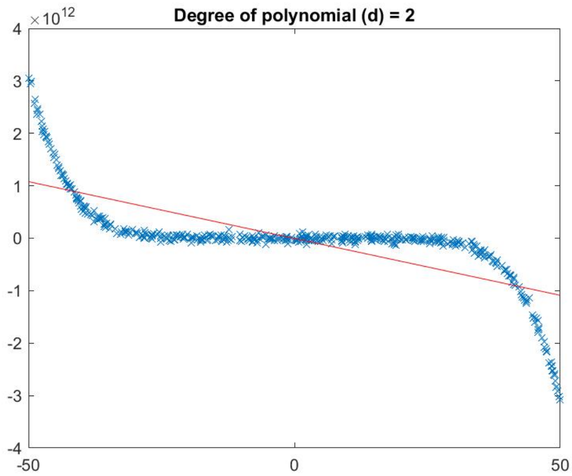
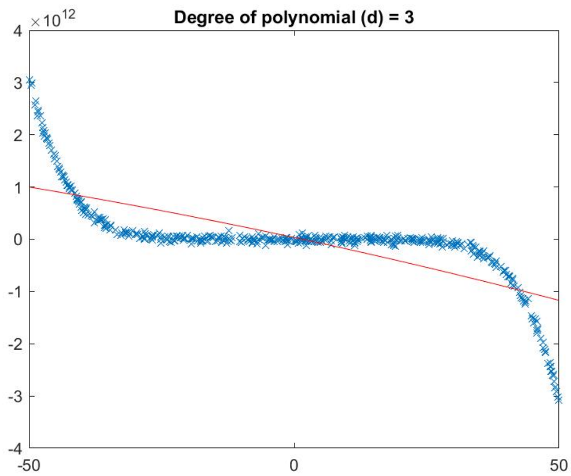
ECE-GY 6143 INTRO TO MACHINE LEARNING

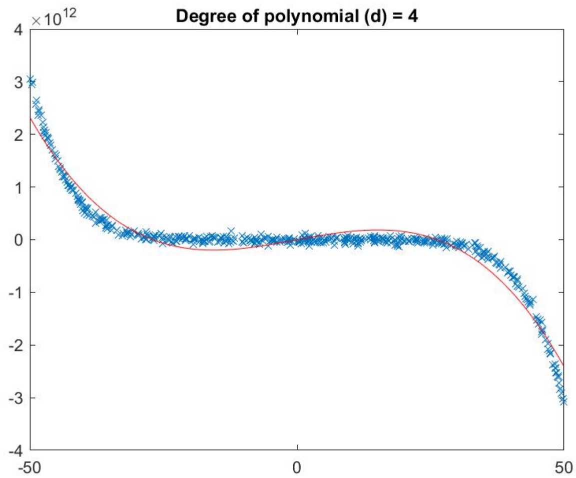
Pratyush Shukla (ps4534)

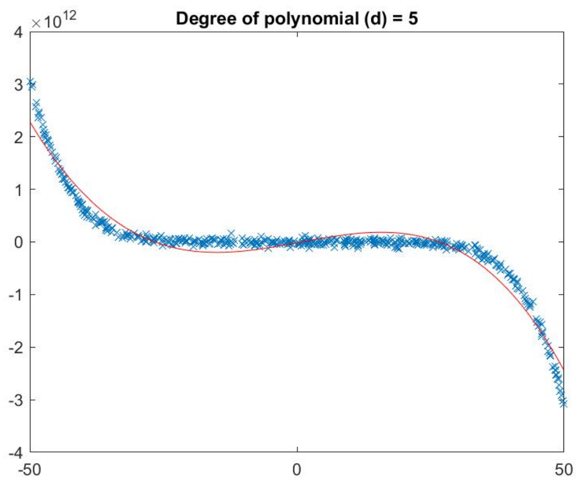
Question 1:

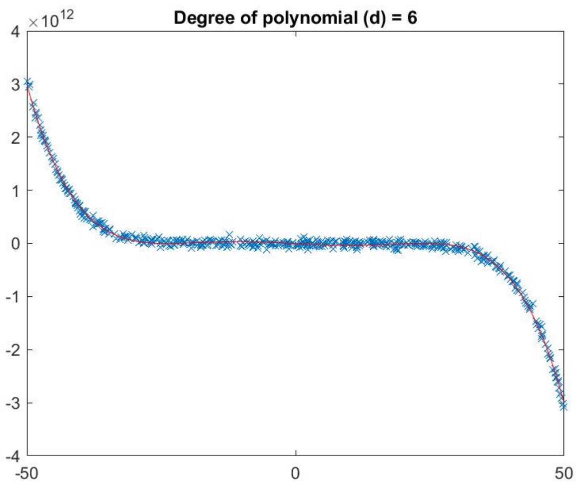


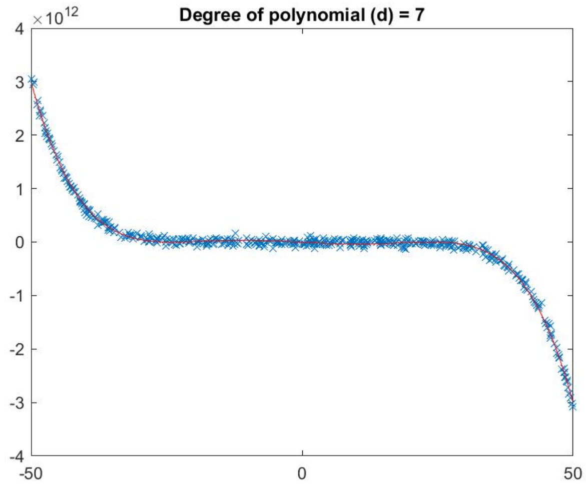


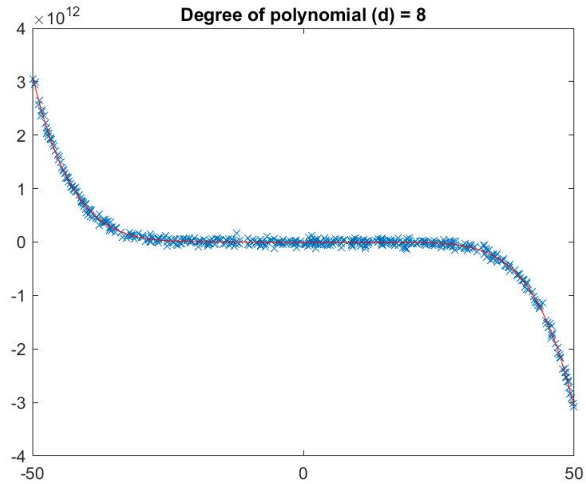


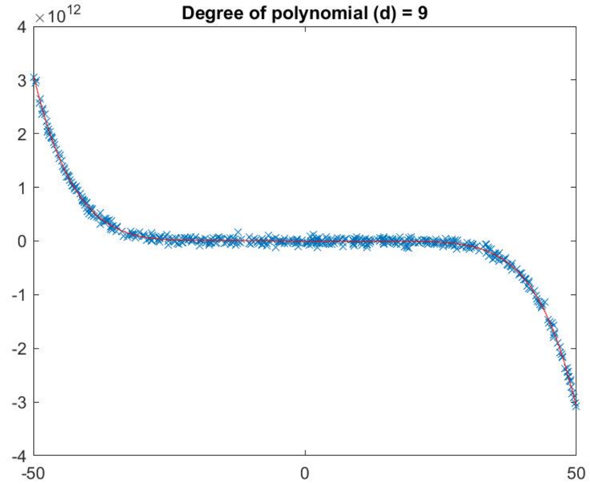


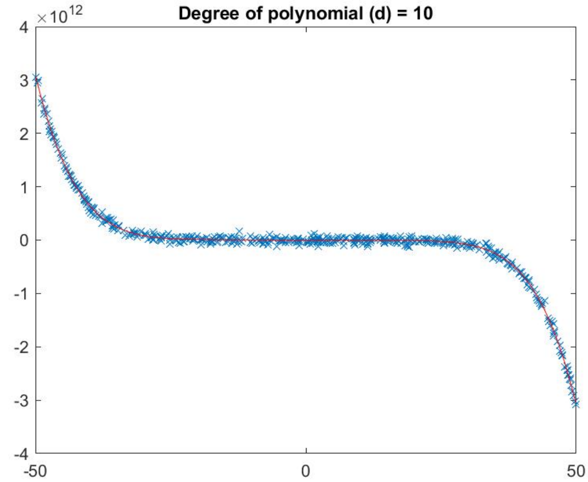






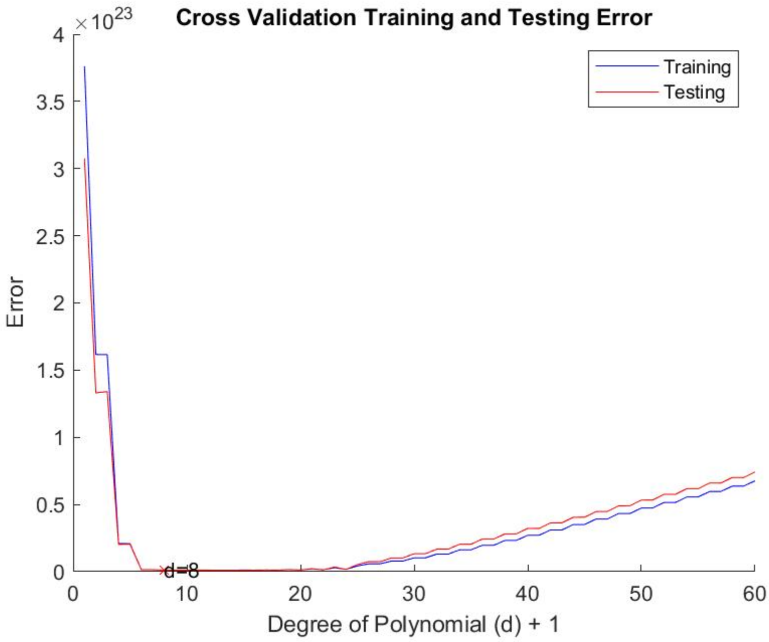






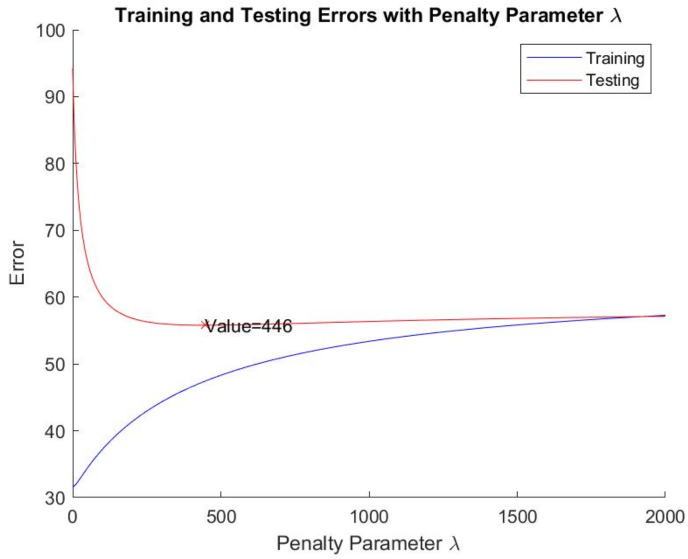
The plots for different values of degree of polynomial d from 1 to 10 is show above. Based on the curve that fits the data points, the most accurate value of d is 5. d above 6 tends to overfit the data.

However, this value of d is not accurate as the data has only been used for training, thus leaving the model out of the assumption that it may have to predict for unseen data points. Hence, the use of cross-validation is to provide the model with training and testing data, divided randomly.



The above plot of training and testing error after cross validation points out that the best value of d i.e., where the errors are least, is 8.

Question 2:



The minimum testing error 55.791615 is obtained for penalty parameter λ 446. We observe that with increase in λ, the training error increases and stabilizes to a certain point while testing error decreases and also stabilizes after a certain point. Note that the value of λ will change for re-runs of the code, due to random splitting of dataset.

Question 3:

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

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

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

$g(-z) = 1 - g(z)$; Hence, proved.

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

Let $g^{-1}(y) = \ln\left(\frac{y}{1-y}\right)$

$$g^{-1}(y) = \ln\left(\frac{y}{1-y}\right) = \ln y - \ln(1-y)$$

$$= \ln\left(\frac{1}{1+e^{-z}}\right) - \ln\left(1 - \frac{1}{1+e^{-z}}\right)$$

$$= \ln 1 - \ln(1+e^{-z}) - \ln\left(\frac{e^{-z}}{1+e^{-z}}\right)$$

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

$g^{-1}(g(z)) = z$; Hence, proved.

Question 4:

The required derivations are as follows:

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

$$R = \frac{1}{N}\sum_{i=1}^{N} (y_i-1)\log(1-f_i) - y_i \log(f_i)$$

To use gradient descent, we minimize R w.r.t θ

Hence, to calculate the gradient –

$$\nabla_\theta R = \nabla_\theta \left[\frac{1}{N}\sum_{i=1}^{N} (y_i-1)\log(1-f_i) - y_i \log(f_i)\right]$$

$$\nabla_\theta R = \frac{1}{N}\sum_{i=1}^{N} (y_i-1)\frac{d}{d\theta}\log(1-f_i) - y_i \frac{d}{d\theta}\log(f_i)$$

$$= \frac{1}{N}\sum_{i=1}^{N} (y_i-1)\frac{d}{d\theta}\left(1-\frac{1}{1+e^{-\theta^T x_i}}\right) - y_i \frac{d}{d\theta}\log\left(\frac{1}{1+e^{-\theta^T x_i}}\right)$$

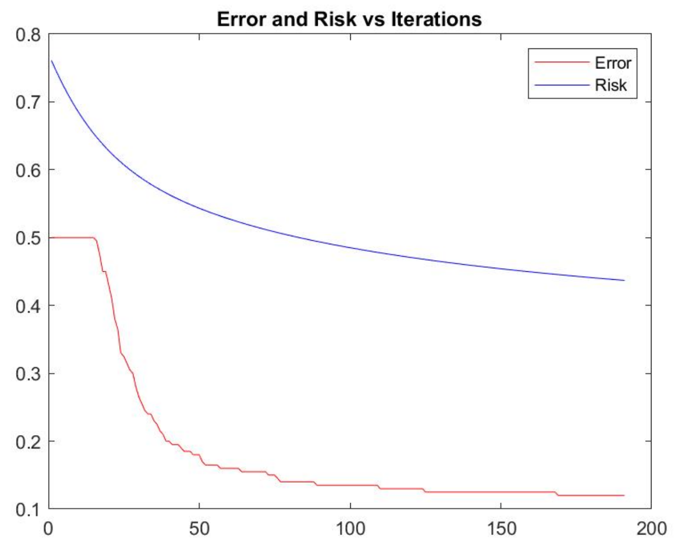
$$= \frac{1}{N}\sum_{i=1}^{N} (y_i-1)\frac{d}{d\theta}\left(\log\left(e^{-\theta^T x_i}\right) - \log\left(1+e^{-\theta^T x_i}\right)\right)$$

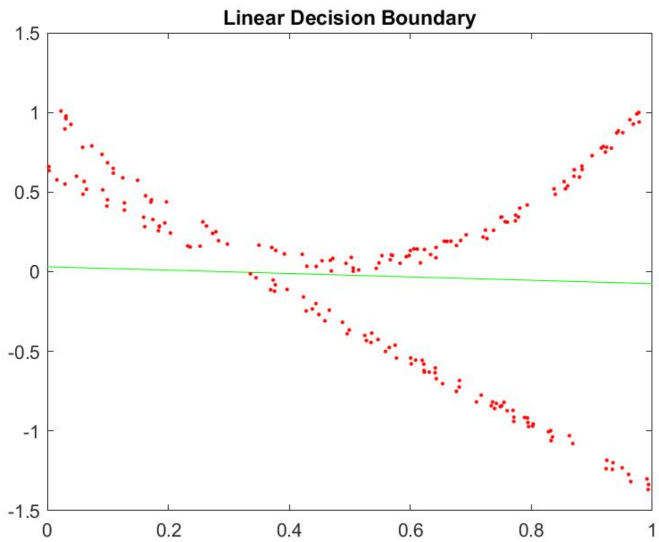
$$- y_i \frac{d}{d\theta}\left(-\log\left(1+e^{-\theta^T x_i}\right)\right)$$

$$\boxed{\nabla_\theta R = \frac{1}{N}\sum_{i=1}^{N}\left[(1-y_i)\left(x_i - \frac{x_i e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}}\right) - y_i\left(\frac{x_i e^{-\theta^T x_i}}{1+e^{-\theta^T x_i}}\right)\right]}$$

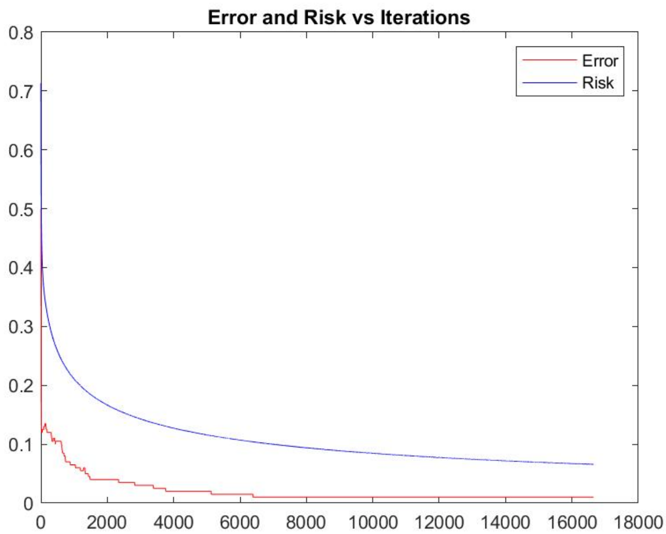
The max number of iterations are 100000.

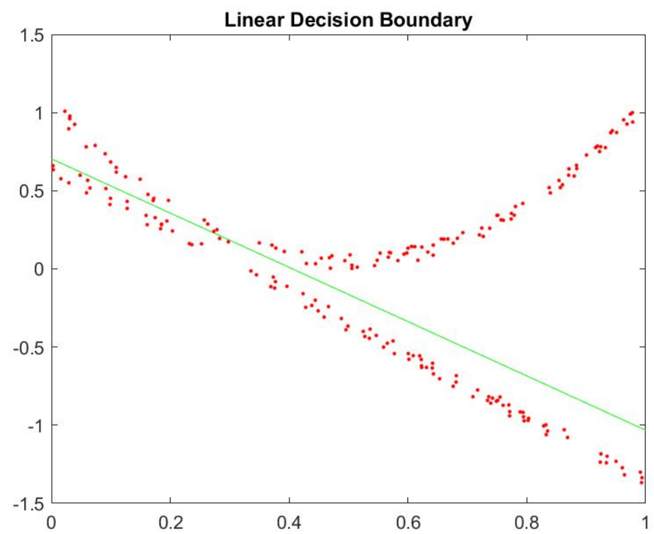
For $\varepsilon = 0.006$ and $\eta = 0.1$, the values of $\theta = [0.2203, 2.0785, -0.0649]$ are obtained at iteration 191 with the following plots:





For ε = 0.001 and η = 0.5, the values of θ = [33.2387, 19.1406, -13.4843] are obtained at iteration 16669 with the following plots:





For ε = 0.001 and η = 5, the values of θ = [101.6543, 52.7610, -38.3416] are obtained at iteration 36866 with the following plots: