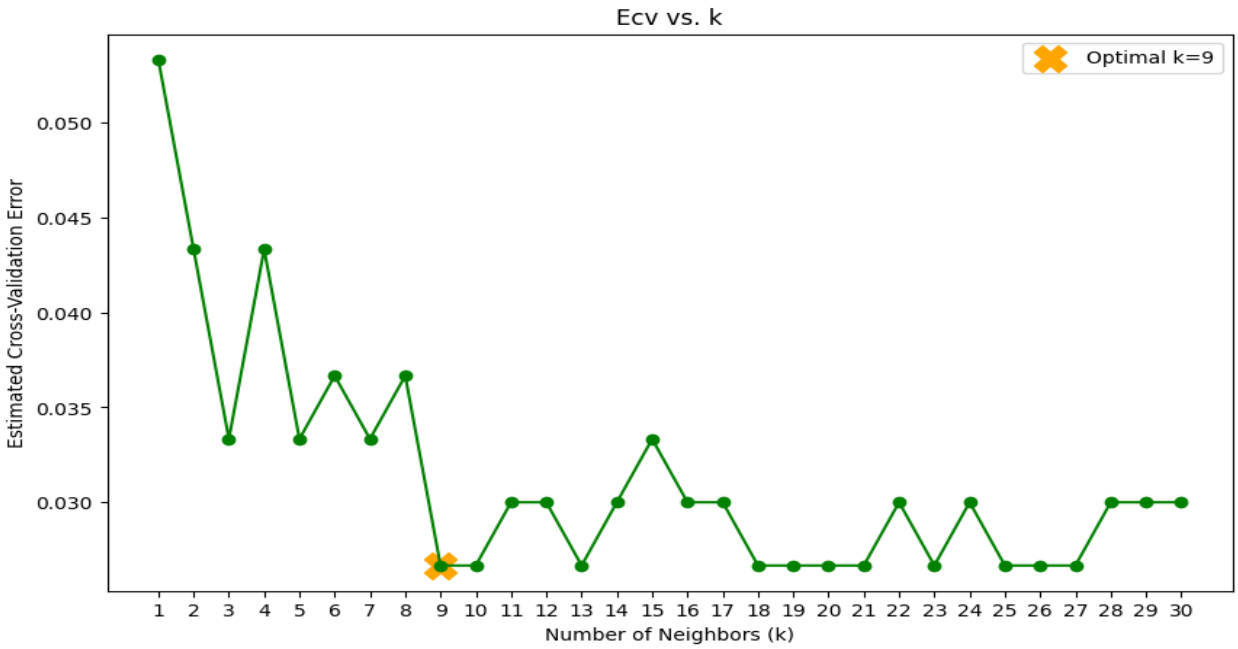
Machine Learning HW 4

Kaushal V. Nerkar

Question 1

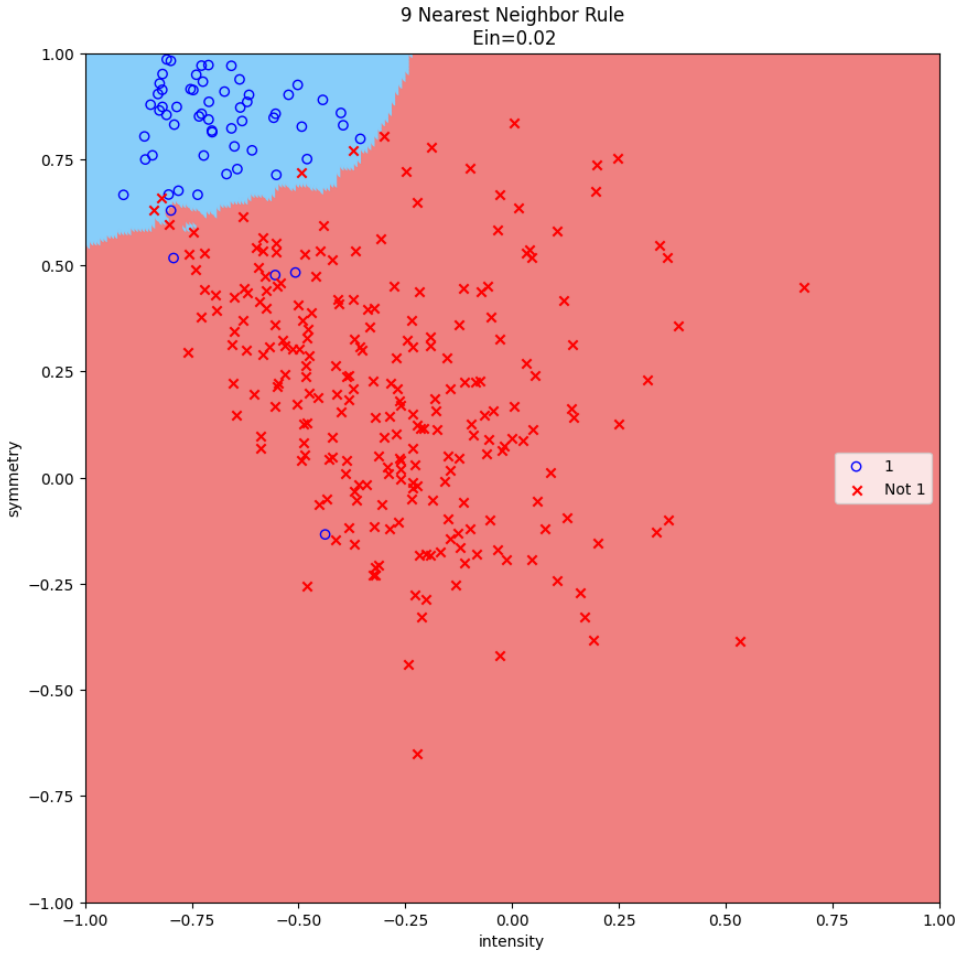
a.



b.

Ecv = 0.02678

Ein = 0.02000



c.

For the optimal k value ‘9’ , E_test is : 0.0318.

Question 2

a. b.

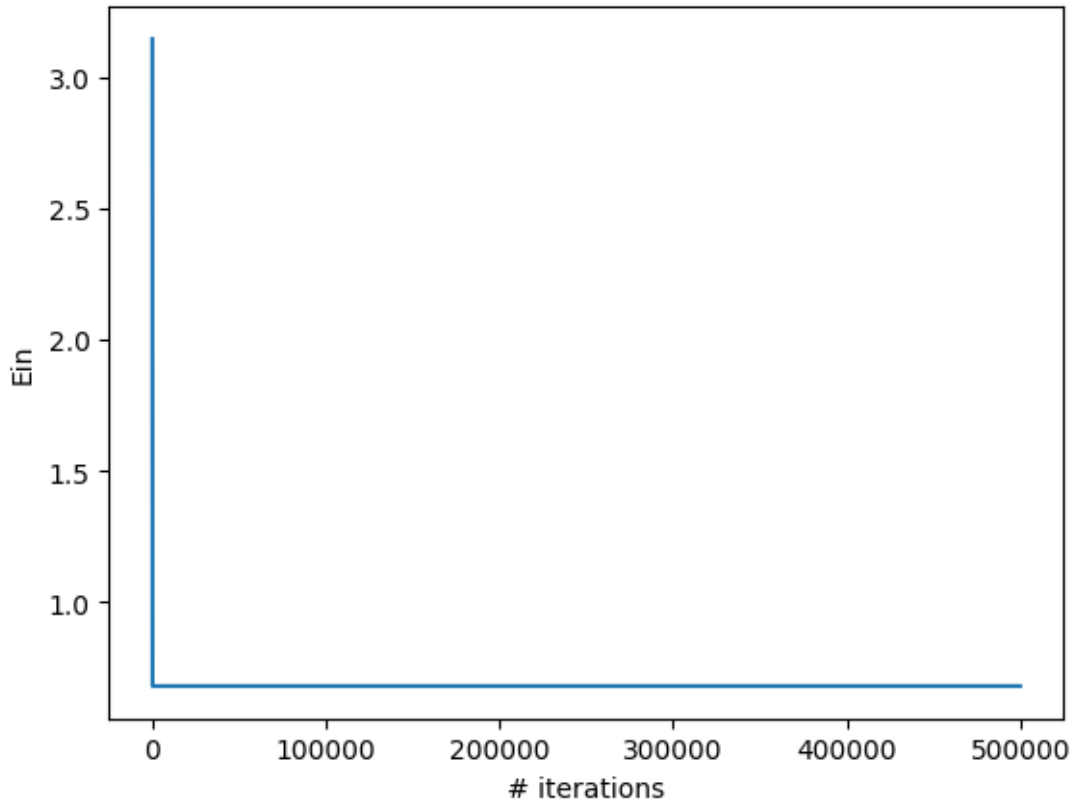
Output for

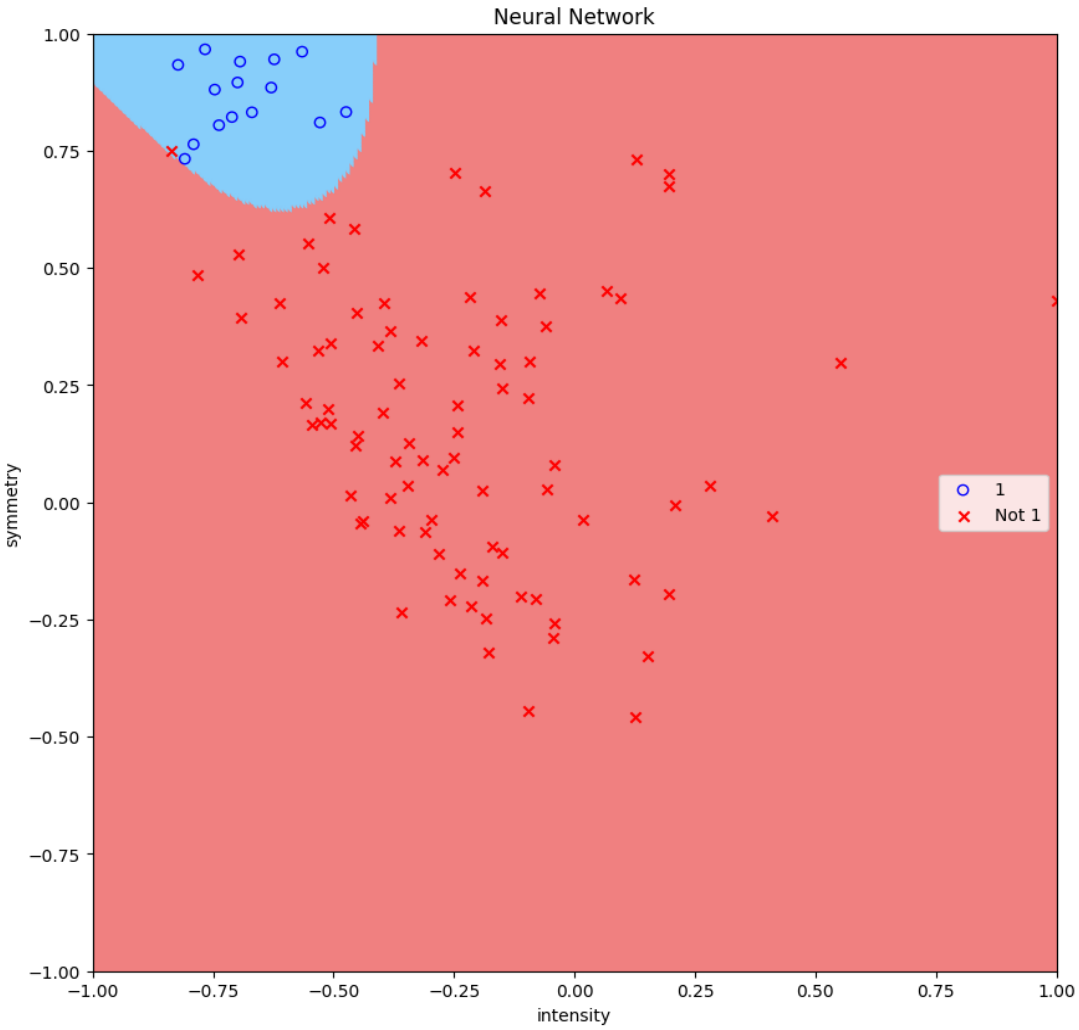
```
Backpropogation
layer 1
[[0.25 0.25]
 [0.25 0.25]
 [0.25 0.25]]
gradient using backpropagation G[1] / ||G[1]||
[[-0.28867513 -0.28867513]
 [-0.28867513 -0.28867513]
 [-0.57735027 -0.57735027]]
layer 2
[[0.25]
 [0.25]
 [0.25]]
gradient using backpropagation G[2] / ||G[2]||
[[-0.68040574]
 [-0.51819303]
 [-0.51819303]]
############################################################


Numerical
layer 1
[[0.25 0.25]
 [0.25 0.25]
 [0.25 0.25]]
gradient using numerical G[1] / ||G[1]||
[[-0.28867531 -0.28867531]
 [-0.28867531 -0.28867531]
 [-0.5773501  -0.5773501 ]]
layer 2
[[0.25]
 [0.25]
 [0.25]]
gradient using numerical G[2] / ||G[2]||
[[-0.68040562]
 [-0.51819311]
 [-0.51819311]]
```

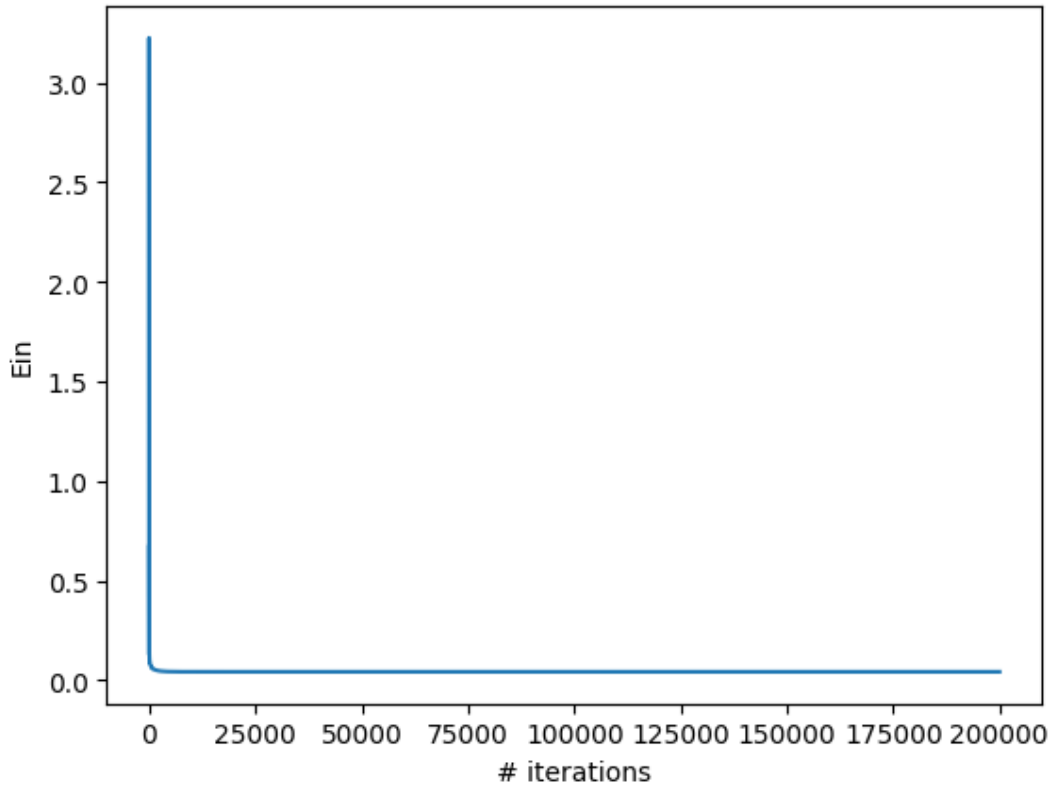
Question 3

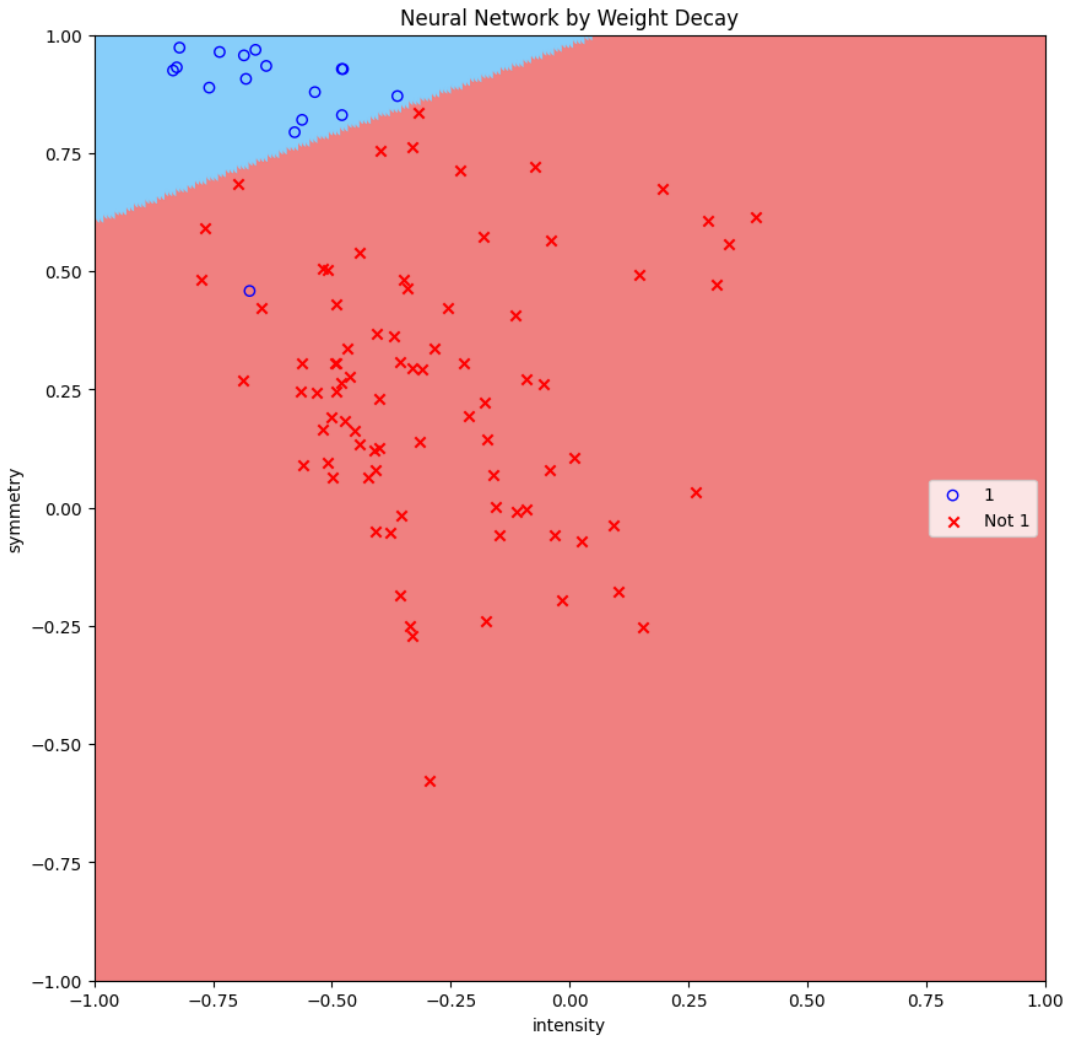
a. Stochastic gradient descent



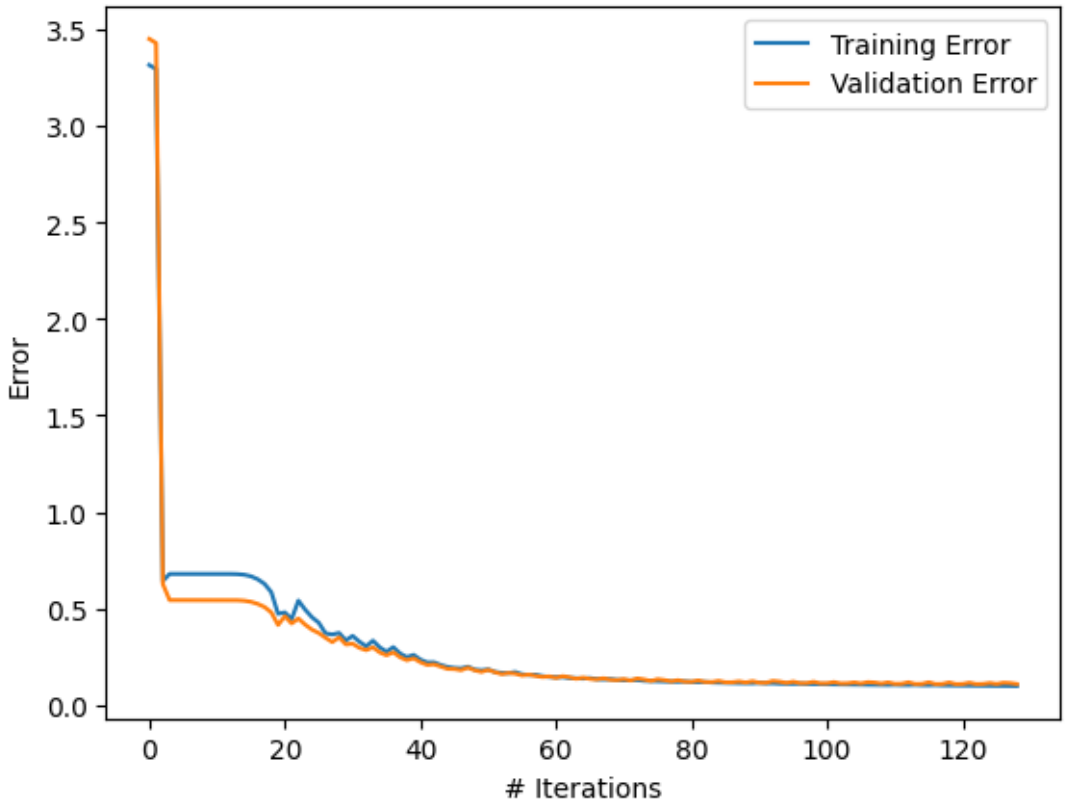


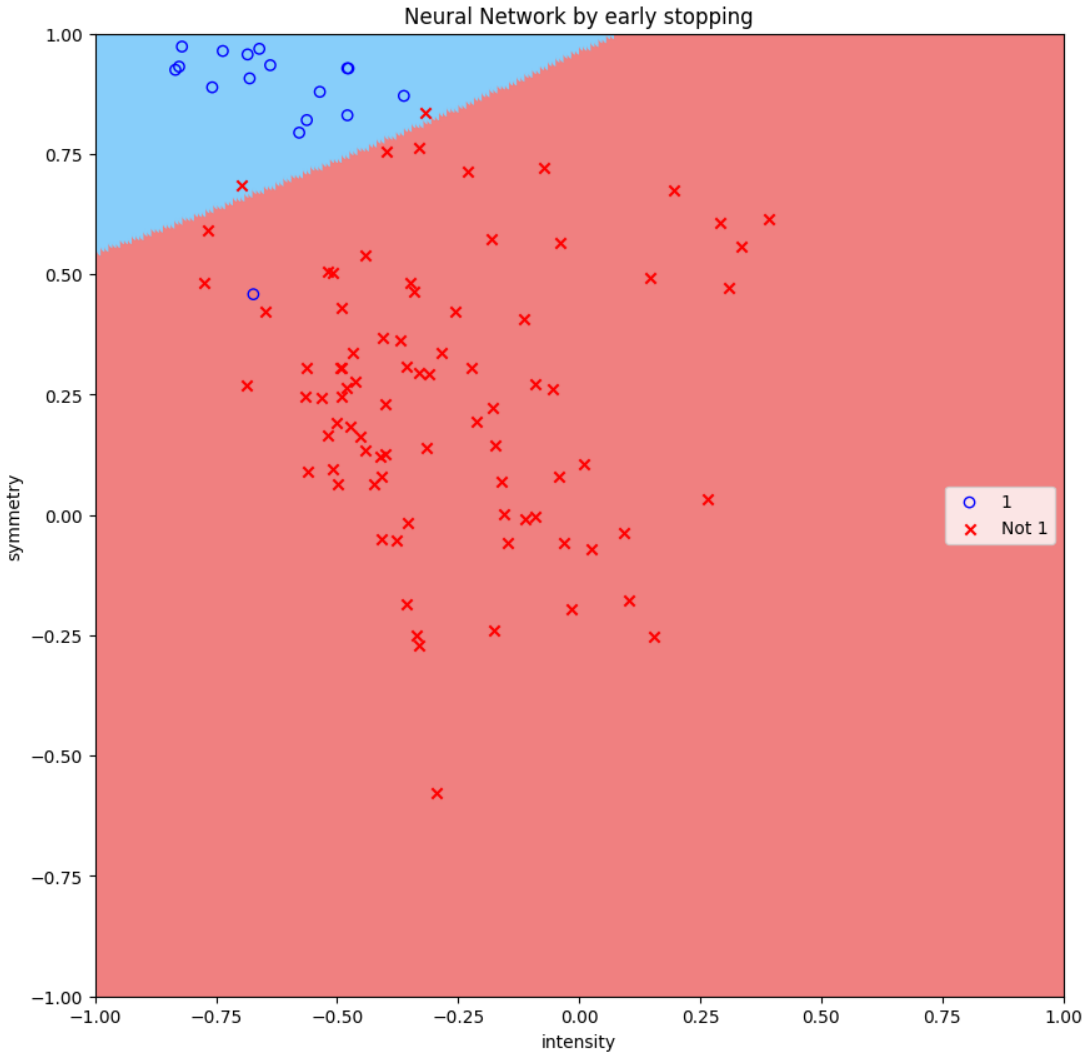
b. Weight decay





c. Early stopping





Question 4

For Linear Regression with 10^{th} order polynomial transform we observe that

optimal lambda found is 1.

And Test Error (E_test) is **0.10752 .**

The test error of 0.10752 suggests that this model performs reasonably well but might struggle with capturing complex nonlinear relationships present in image data.

The relatively higher error indicates that the linear regression model with a 10th-order polynomial might not be fully capturing the intricacies of distinguishing digit 1 from others in the MNIST dataset.

For K nearest Neighbor rule, we observe that 'k' optimal is 9.

And Test Error (E_test) is **0.0318.**

The k-NN model exhibits a significantly lower test error of 0.0318, indicating superior performance in distinguishing between digit 1 and other digits in the dataset.

Lower test error suggests that the k-NN approach with an optimal k value of 9 is more effective in capturing the complex patterns present in the image data compared to the linear regression model.

For Neural Network with early stopping

We get the number of iterations that are executed in neural network are 125 (after that weights are not getting that optimized so it stops early saving some computation time).

Test Error (E_test) is **0.10354**

The neural network with early stopping has a test error of 0.10354, which is lower than the linear regression model but slightly higher than k-NN.

The neural network, despite stopping early at 125 iterations, showcases competitive performance in distinguishing between digit 1 and other digits in the MNIST dataset.

Optimal for lambda (hyperparameter for linear regression, k for k-NN) greatly influences model performance. Further tuning it might improve the performance of the model.

Link to collabs :

https://colab.research.google.com/drive/1miPO2KE-eVX-FsY4esvWVbG8LABiVR7z?authuser=1#scrollTo=xIvrNuvBTSac&uniqifier=3

https://colab.research.google.com/drive/1_RNSCBEX6Ccmk0RCXO29XE5x_LXwidIr?authuser=1#scrollTo=6ehnrBOcAUea

https://colab.research.google.com/drive/1ve8t6YnAzQC71S4wk94FMTRqFgPZPbMA?authuser=1#scrollTo=73736df9