Question 5

5 a:

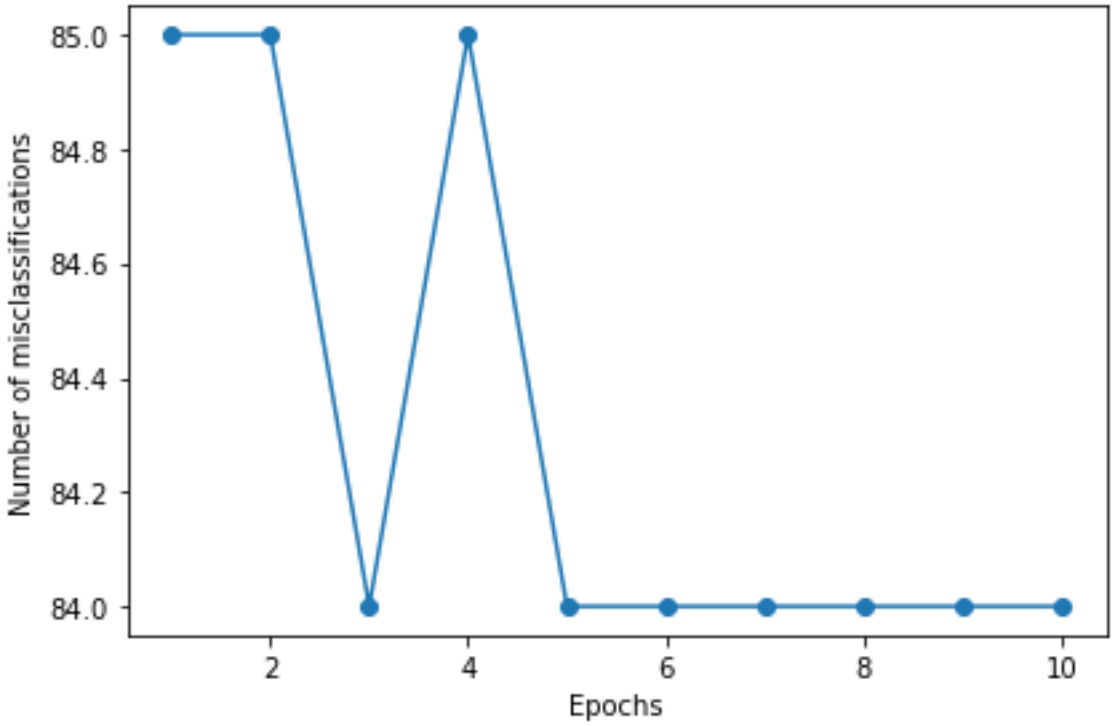
Perceptron accuracy 0.18

Adaline accuracy 0.5

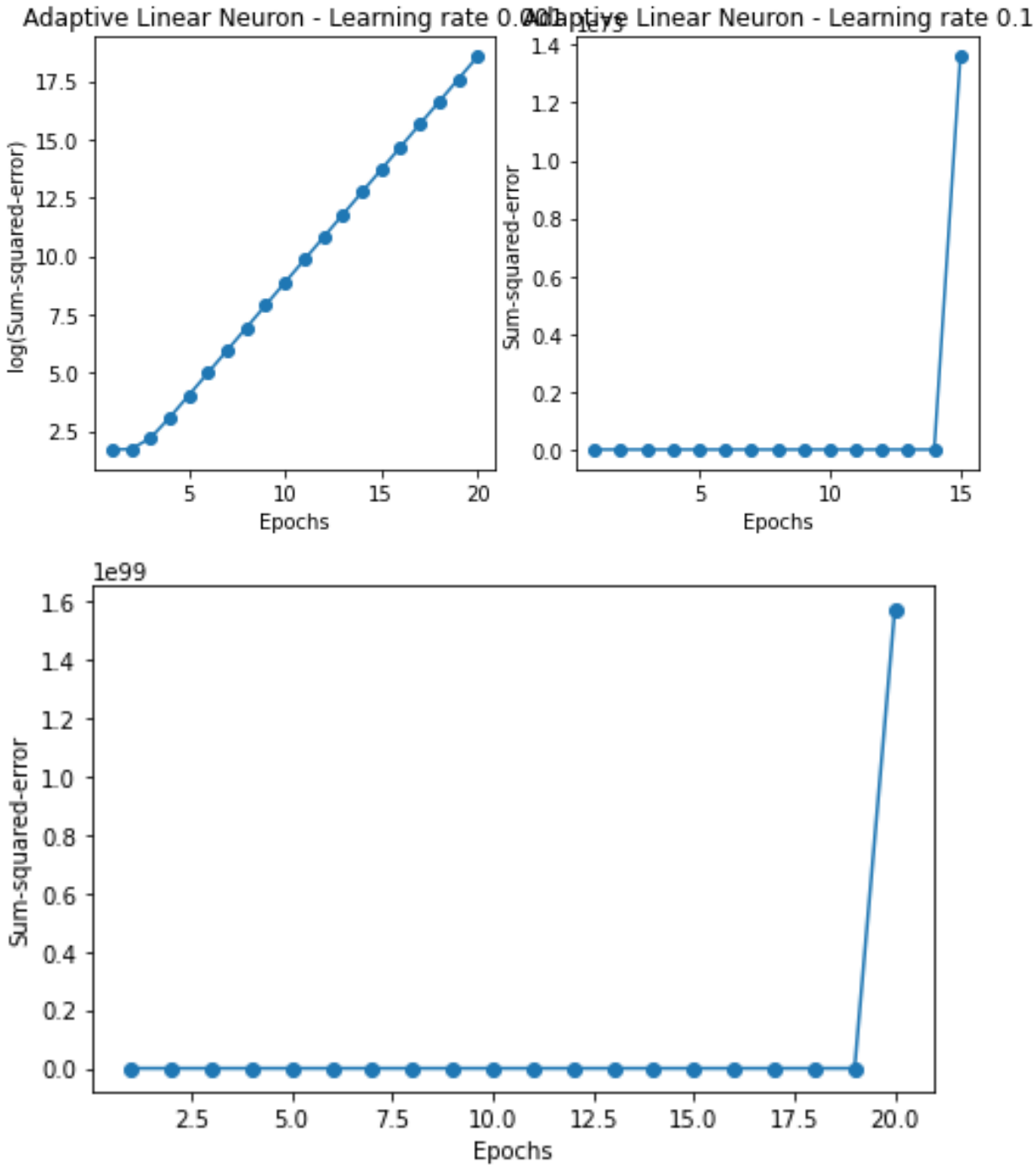
SGD  accuracy 0.9933333333333333

5 b. Perceptron

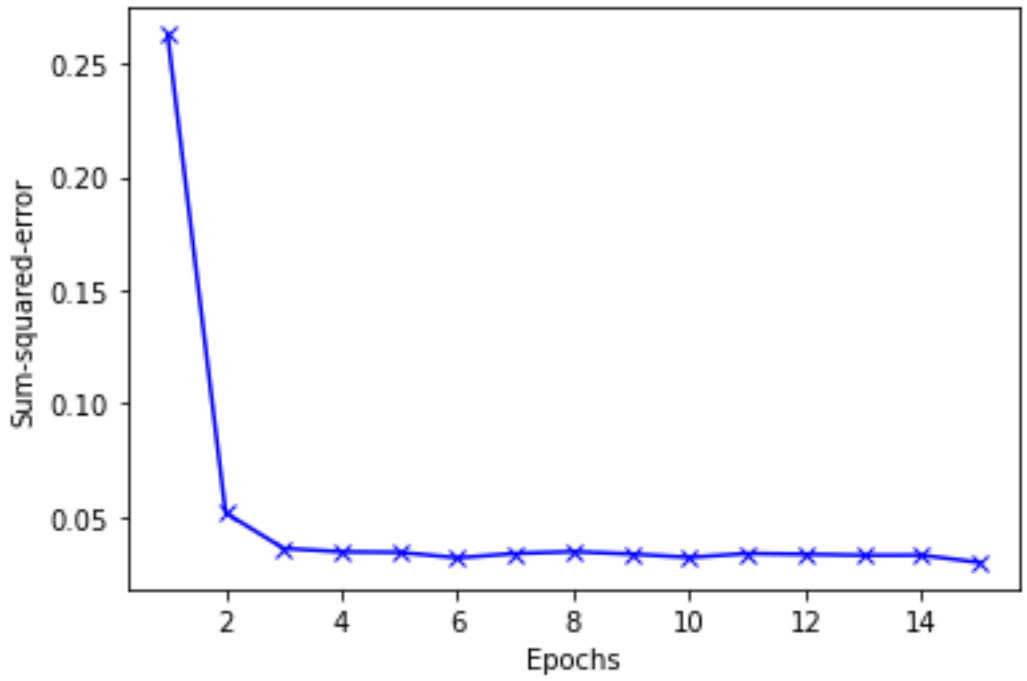
Errors [85, 85, 84, 85, 84, 84, 84, 84, 84, 84]



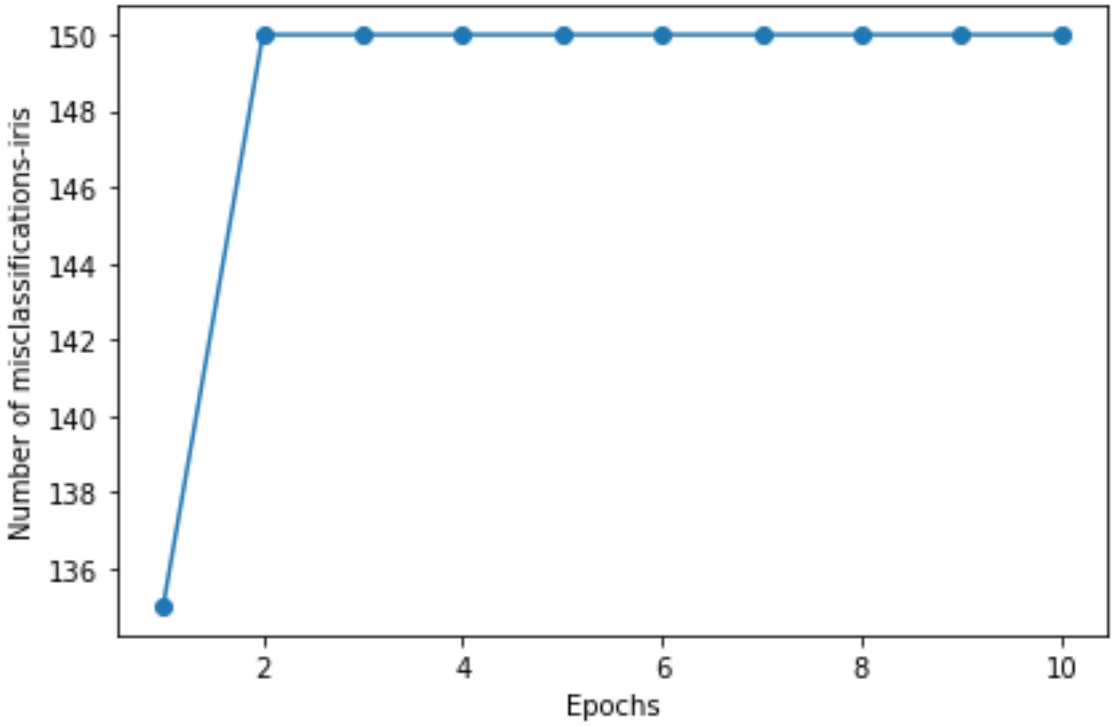

cost for adaline [50.0, 237671.0570255, 38038116604.39305, 6208335248252397.0, 1.0132966259212845e+21, 1.6538573042825e+26, 2.699351712983616e+31, 4.4057607941869775e+36, 7.190885160400322e+41, 1.1736640232100418e+47, 1.9156017773212438e+52, 3.1265593020734516e+57, 5.103029859917859e+62, 8.328936455465163e+67, 1.3594116511850714e+73, 2.2187707245204487e+78, 3.6213780599109967e+83, 5.910650842772045e+88, 9.647099200137227e+93, 1.5745562621263036e+99]

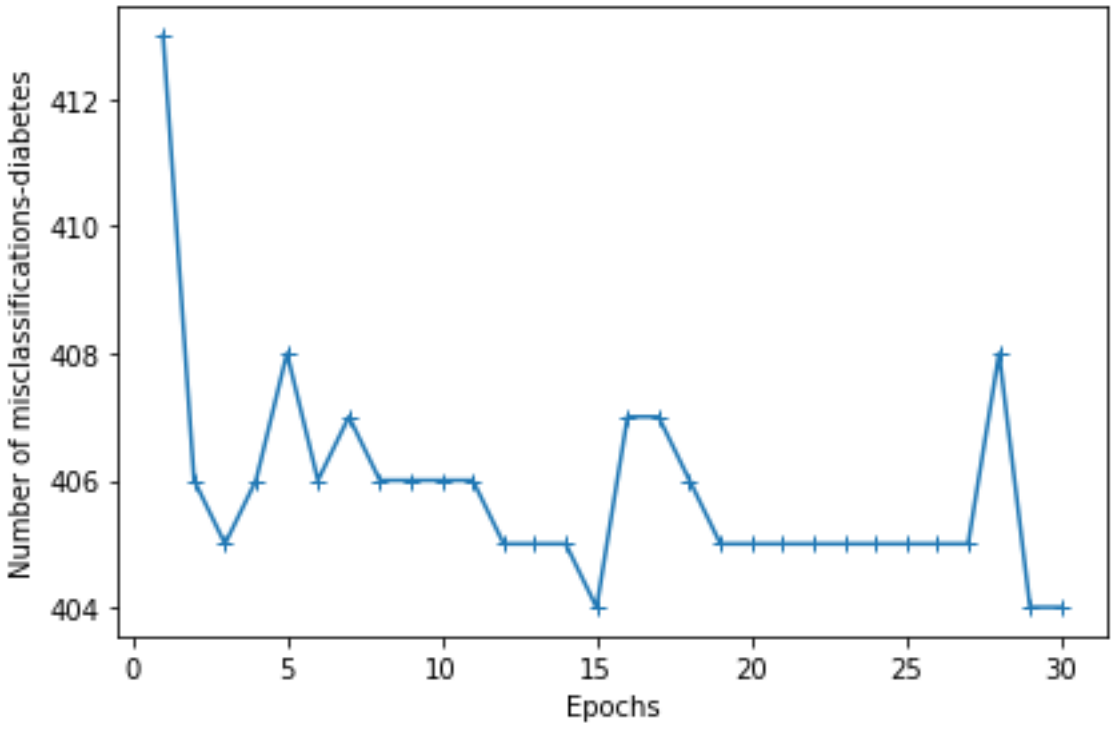


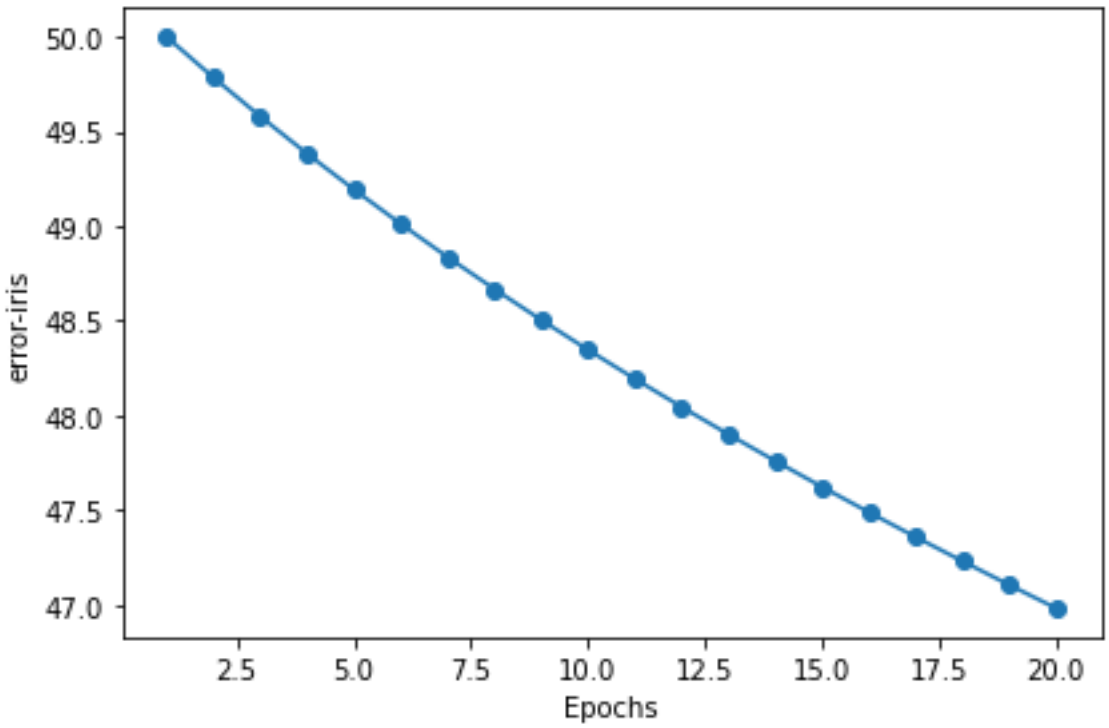

cost of SGD [0.26299001261581995, 0.051714396826822336, 0.03603994927276267, 0.03452658149463849, 0.03431528990451994, 0.031881275466748134, 0.033823730961893256, 0.03461866708661809, 0.03355214574135058, 0.032076625590373296, 0.03379813414958148, 0.033454502997371155, 0.0330083667559797, 0.03304139830549488, 0.02980135150824148]

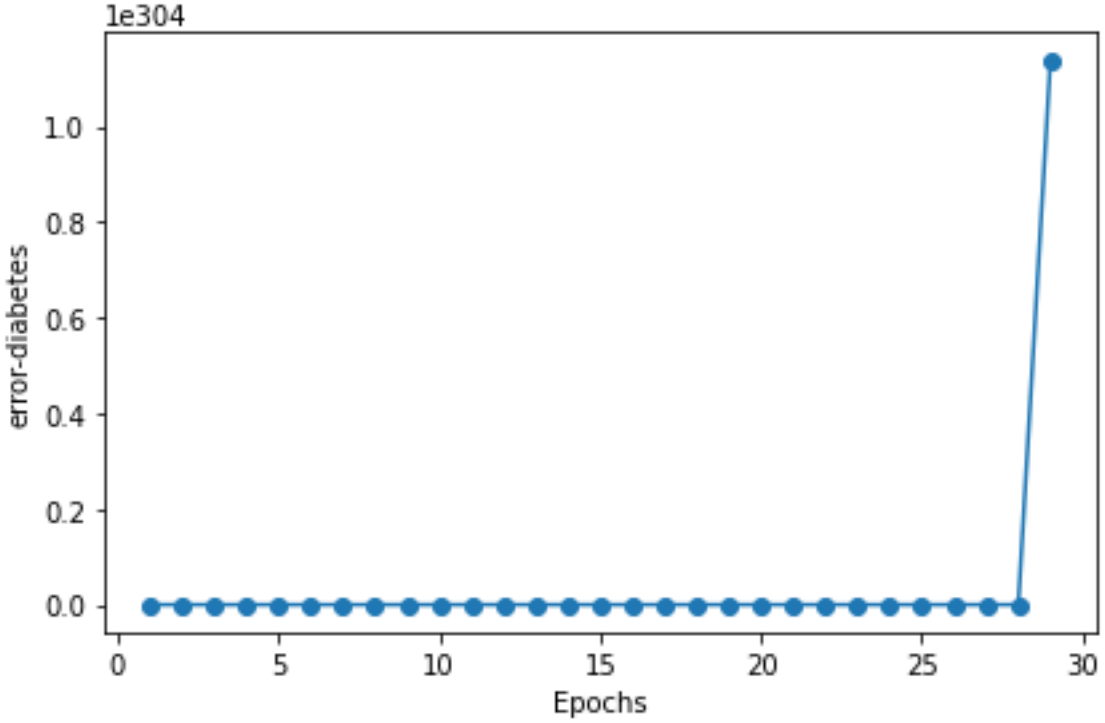

0.25
0.20
0.15
0.10
0.05
Sum-squared-error
2
4
6
8
10
12
14
Epochs

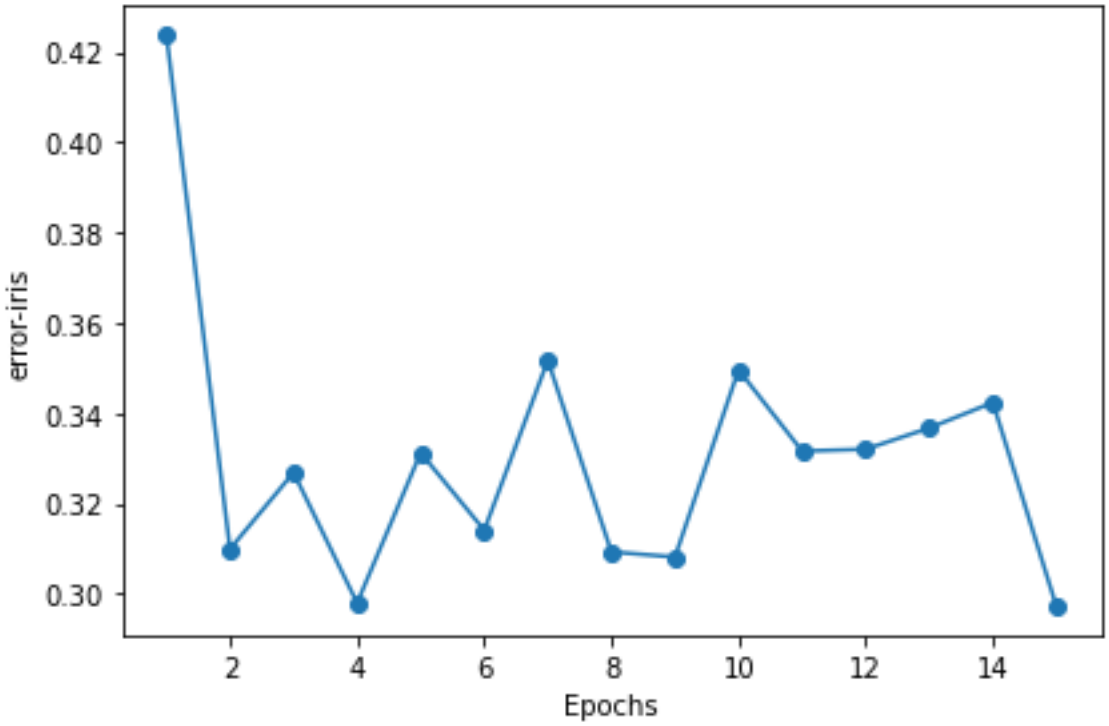
Question 5 c

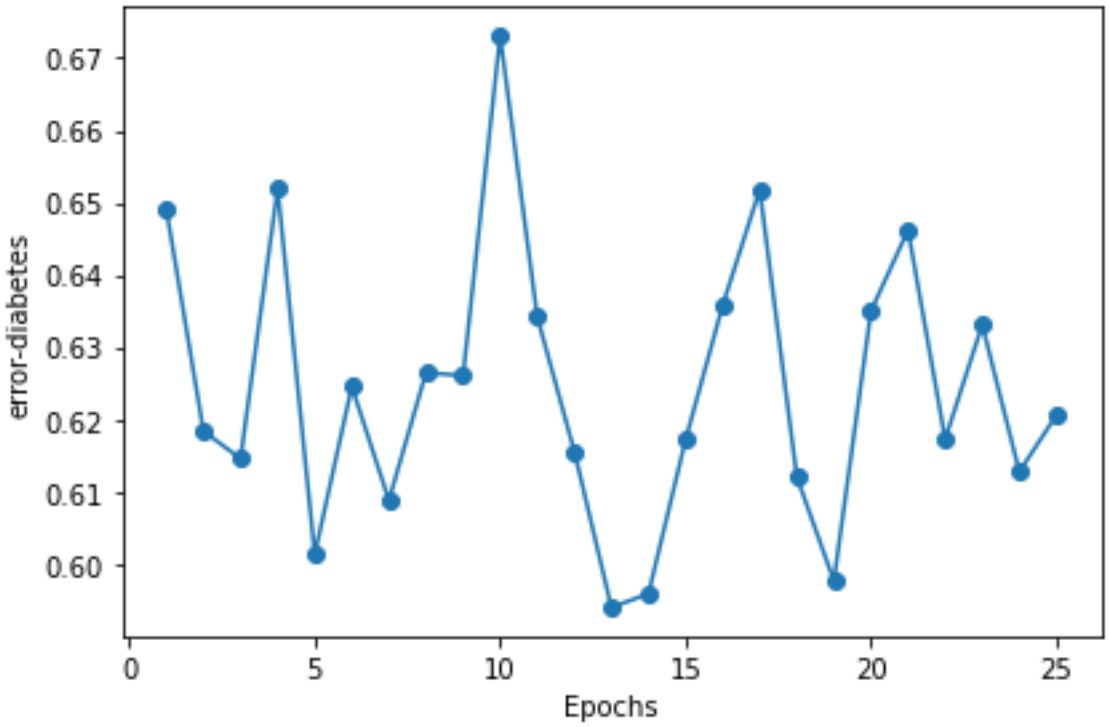

Number of misclassifications-iris
150
148
146
144
142
140
138
136
2
4
6
8
10
Epochs


Number of misclassifications-diabetes
412
410
408
406
404
0
5
10
15
20
25
30
Epochs


50.0
49.5
49.0
48.5
48.0
47.5
47.0
error-iris
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epochs


1e304
1.0
0.8
0.6
0.4
0.2
0.0
error-diabetes
0
5
10
15
20
25
30
Epochs


error-iris
0.42
0.40
0.38
0.36
0.34
0.32
0.30
2
4
6
8
10
12
14
Epochs


error-diabetes
0.67
0.66
0.65
0.64
0.63
0.62
0.61
0.60
0
5
10
15
20
25
Epochs

Question 5 d:

Feature scaling is a technique of standardizing the features in the data in a fixed range. When the data consists of features of varying magnitude, units and ranges this technique will be used. With feature scaling accuracy can be improved.

Min-Max scaling:

The data will be rescaled from original range so the values will be in range of 0 and 1 in normalization.

In normalization you should be able to estimate the minimum and maximum observable values. You can estimate the values from available data.

Values will be normalized by y= (x-min)/(max-min)

When X is outside of bounds of minimum and maximum values the resulting value will not be in range od 0 and 1.

Z-score normalization:

It address the problem of outliers without requiring prior knowledge of what the reasonable range is by linearly scaling the input using the mean and standard deviation estimated over the training dataset.

Y= (x-mean)/standard deviation

With feature scaling accuracy will be improved.

https://vitalflux.com/feature-scaling-stratification-model-performance-python/

https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/

Question 5 e:

Batch size is one of the important hyperparameter to control modern deep learning systems. Often usage of large batch size to train the model allows computational speedup. However large batch size will lead to poor generalization. When we try to convex functions to optimize there will inheritance war between small and big batch size. When you use the batch size equal to dataset can guarantee convergence of global optima. If you use small batch size will have faster convergence to good solutions. This explains small batch size allow the model to learn before seeing the data. But there is no guarantee of converge to global optima when you use small batch size. If you start with a small batch size there will be faster training and benefits of guaranteed convergence.

https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e