Group members- MD Rehan Zeya 2014csb1059
- Viren Gupta 2014csb1080

Q1

iv) As we increase the value of H , the training error first decreases and then increases.

As the number of epochs increases from 0, the change in error for the highest value of H is maximum. It means higher the value of H, more faster it will converge to its minima. So at very low values of epochs, high value of H gives least training error while at high values of epochs, higher H give maximum error.

The plots shows that from H=2 to H=4, the training error decreases but at H=64, training error is maximum.Because the plot shows that first training error decreases upto a value of H and then increases.

v)

sd=0.75

1000 epochs - Training error - 13%
5000 epochs - Training error - 11.67%
10000 epochs - Training error - 10.67%

Training error decreases as we increase the number of epochs.

As we increase the number of epochs ,training error decreases.So here the model tries to overfit the training instances to reduce its training error as number of epochs increases. The order of the polynomial which represents the decision boundary increases due to overfitting. We can see that here it changes from almost linear to some higher order.

Q2

Approach:

We used 5 cross validation process to find the optimum value of number of hidden layer nodes and learning rate. At each k-fold iteration, we divided the given data into 80% training and 20% test data. The test data are mutually exclusive in each iteration of k-fold process. Then we standardized the training data and standarize the test data using the mean and variance of training data. Then we performed forward and backward pass on it. We set the number of epochs=1000. We update the parameters with a batchsize =10. We vary values of H and eta and find those values of H and eta for which average test error over k-fold is minimized. The plot for this is also attached in the submission.

Once we found the optimum parameters, we re-trained our model on the training set and found out the optimum values of w and v parameters. With these values we calculated the test error on the given testgenredata.mat. NOTE: Before predicting the class of the intances of testgenredata.mat first standarize this test data using mean and variance of the training data and then find the prediction .

I am attaching w and v for each k fold iteration on my model with optimal parameters(H=64 and eta=0.008).

Then use each pair of w and v to calculate the prediction of test data.We will get five prediction of our test data from five pairs of w and v. Then the final prediction of our test data will be the mode value of all the five predictions.(e.g. if instance i is classified as 1 , 3, 3, 4, 2 with each pair of w and v then the final prediction of ith instance is 3 as it is maximum among the five prediction).

NOTE:Just run the code check.m to obtain the prediction of testgenredata.mat.Make sure all the w's and v's are present in the same folder.