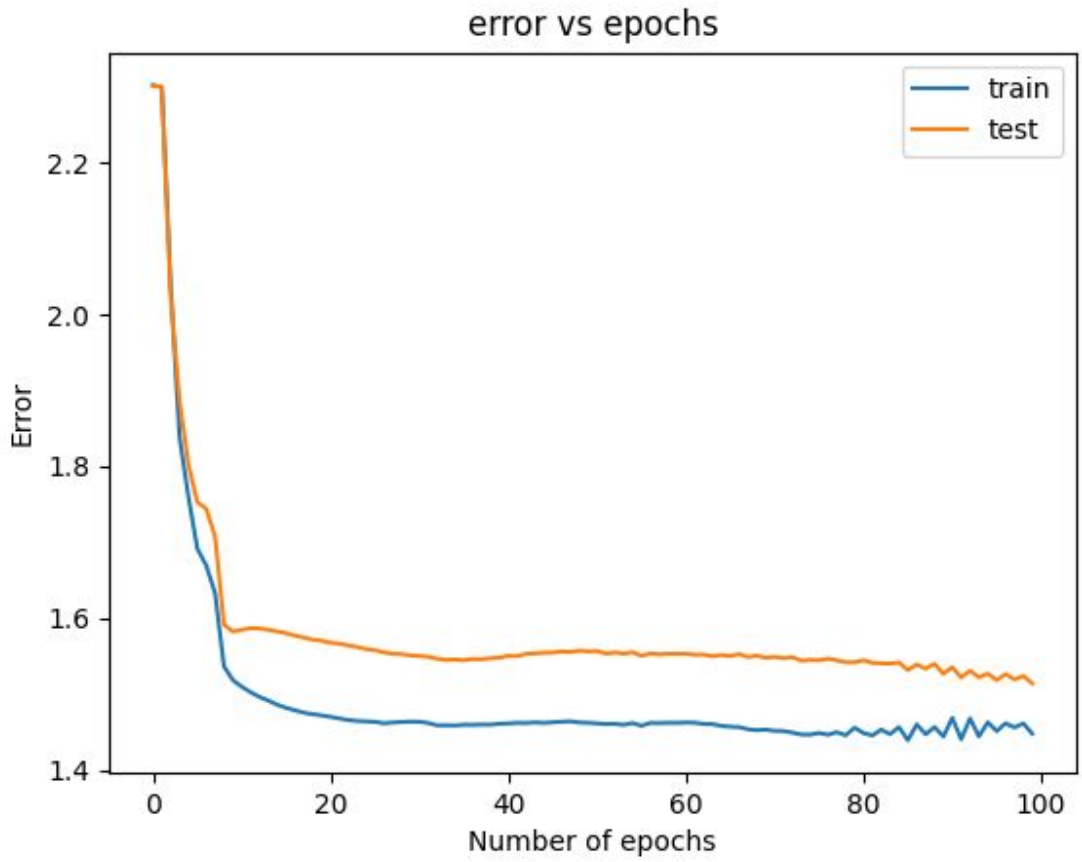
ML Assignment-3

Akshala Bhatnagar
2018012

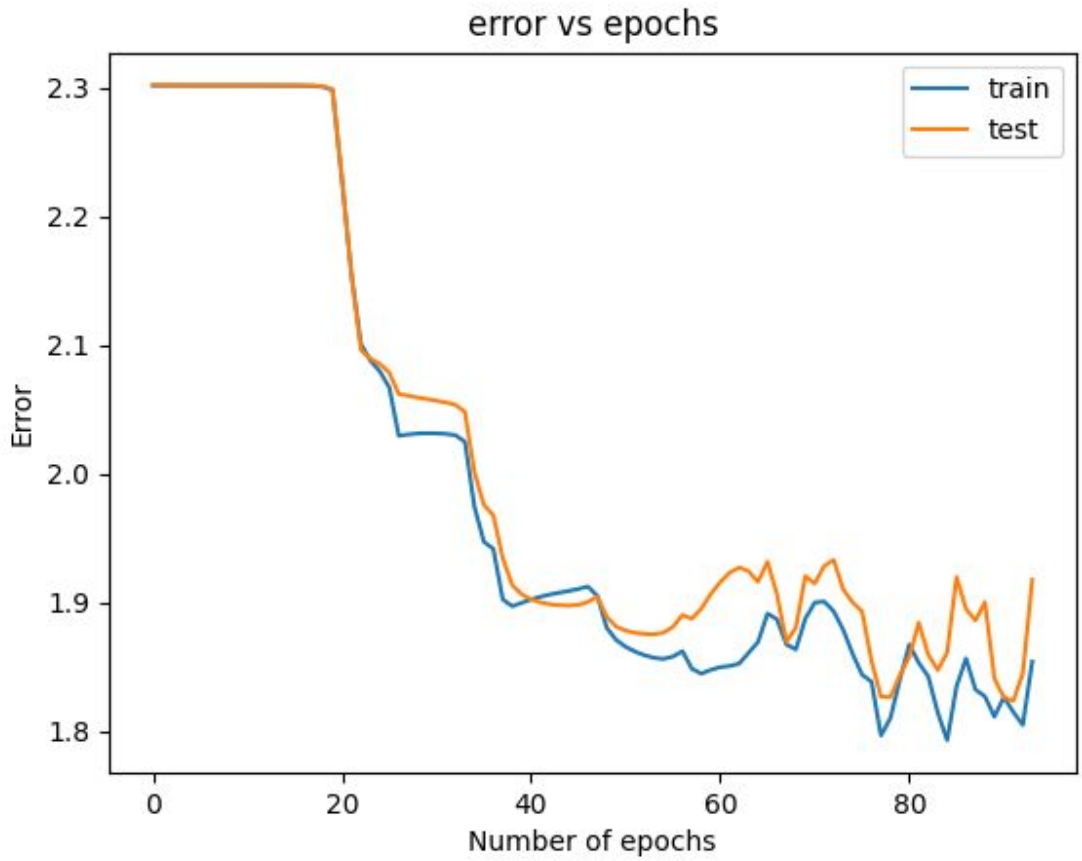
1. Implemented neural network
2. 1)

Activation Function	Test Accuracy	Train Accuracy
ReLu	0.9778	0.9925166666666667
sigmoid	0.737	0.7340166666666667
linear	0.8534	0.85395
tanh	0.9613	0.9708833333333333

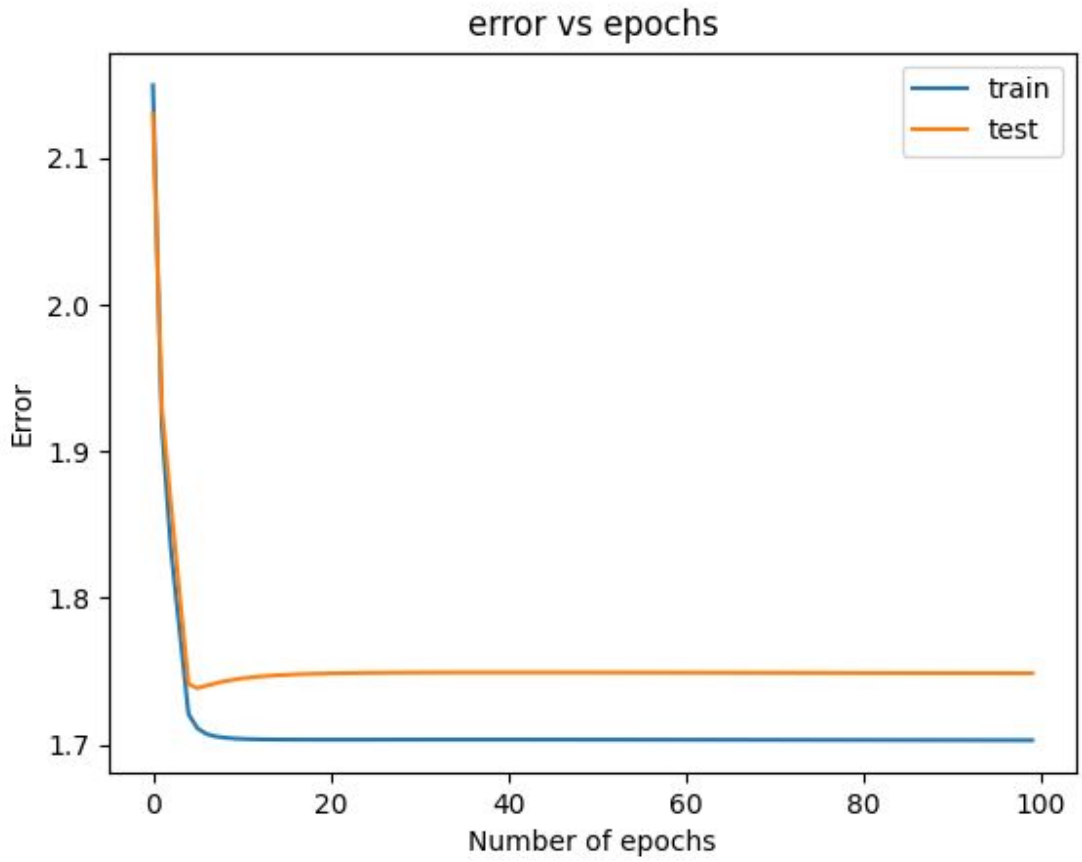
2) **ReLU**


error vs epochs
train
test
Error
2.2
2.0
1.8
1.6
1.4
0
20
40
60
80
100
Number of epochs

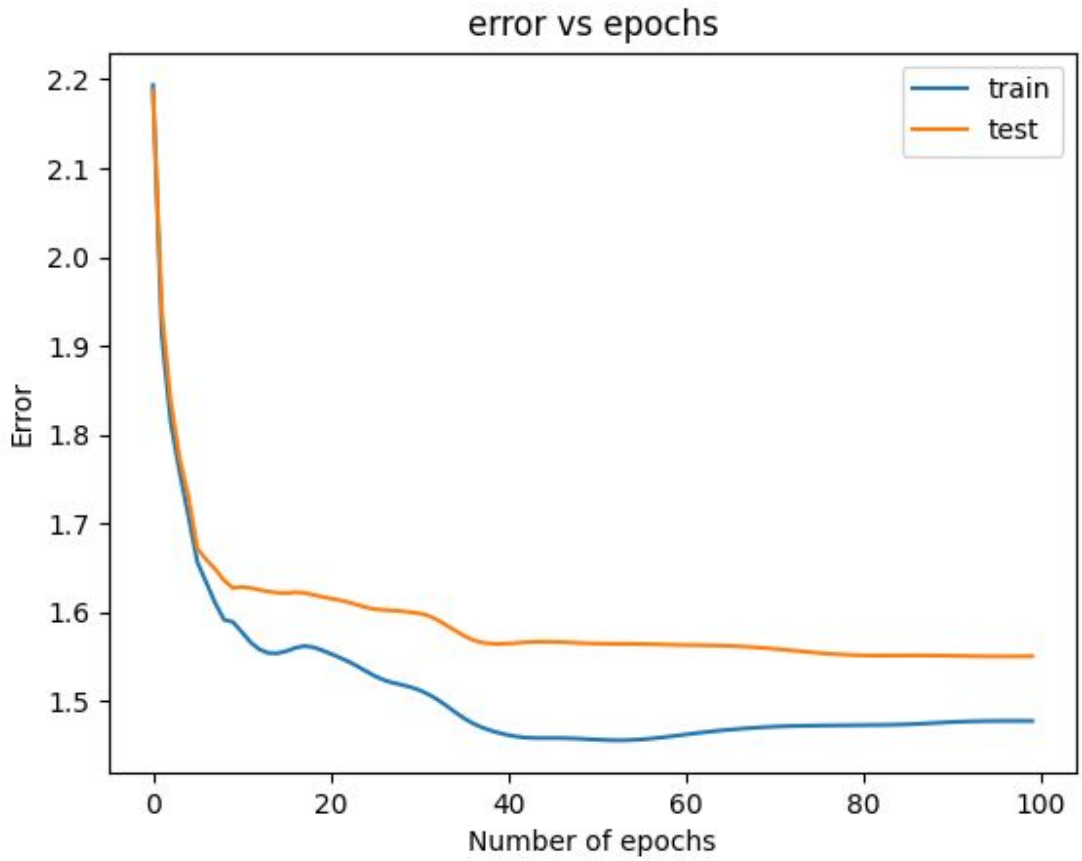
Sigmoid


error vs epochs
train
test
Error
2.3
2.2
2.1
2.0
1.9
1.8
0
20
40
60
80
Number of epochs

Linear


error vs epochs
train
test
2.1
2.0
1.9
1.8
1.7
Error
0
20
40
60
80
100
Number of epochs

Tanh

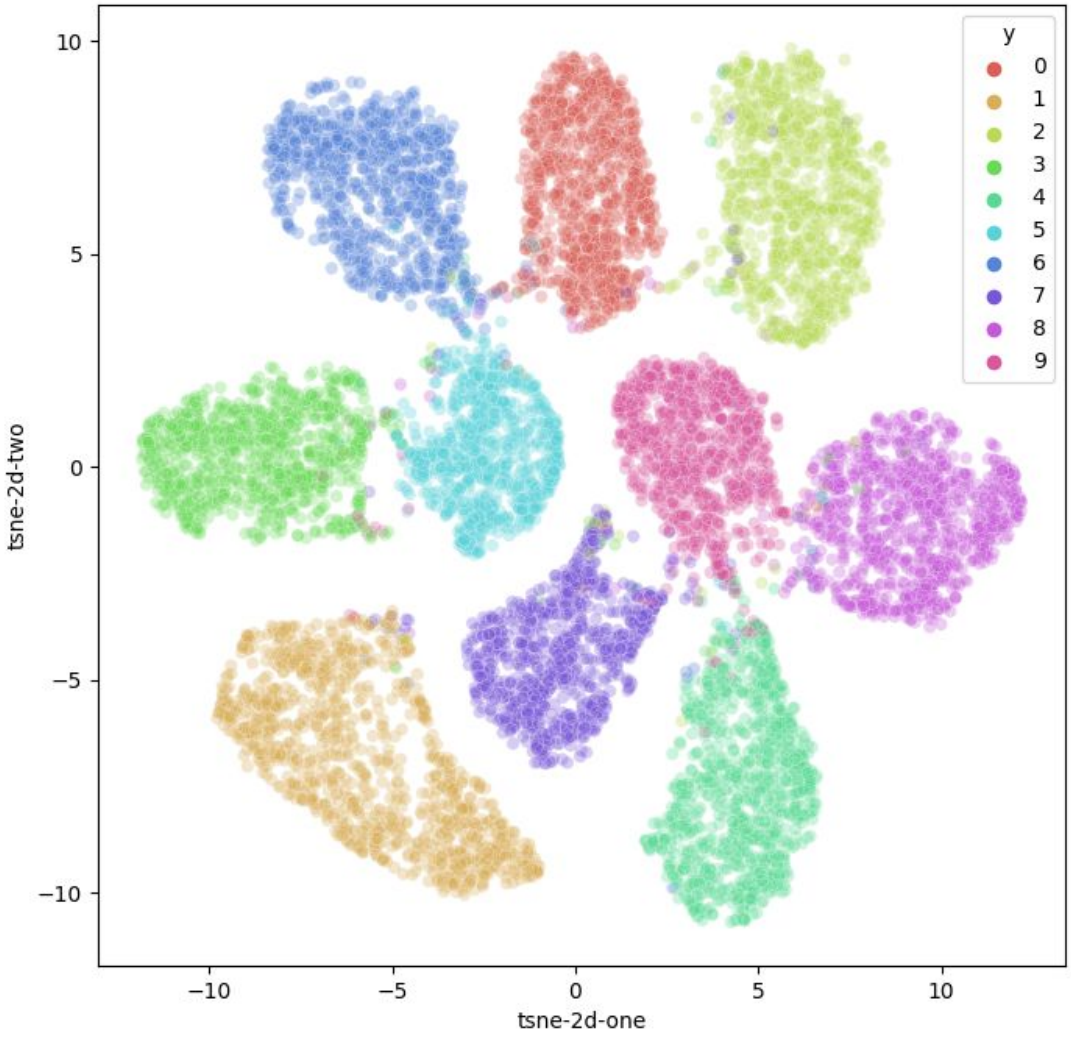


3) For every case the activation function for the output layer should be softmax. The output consists of values 0-9. Softmax function converts the weighted sum values into probabilities that will add up to 1, ie- the scores are converted into a normalized probability distribution. Now, the class corresponding to the highest probability can be taken as the predicted class. None of the other activation functions given perform this task.

4) Total number of layers = 4

Number of hidden layers = 3

5) Best test accuracy is obtained for ReLu. This is the tsne plot for features of the last hidden layer.



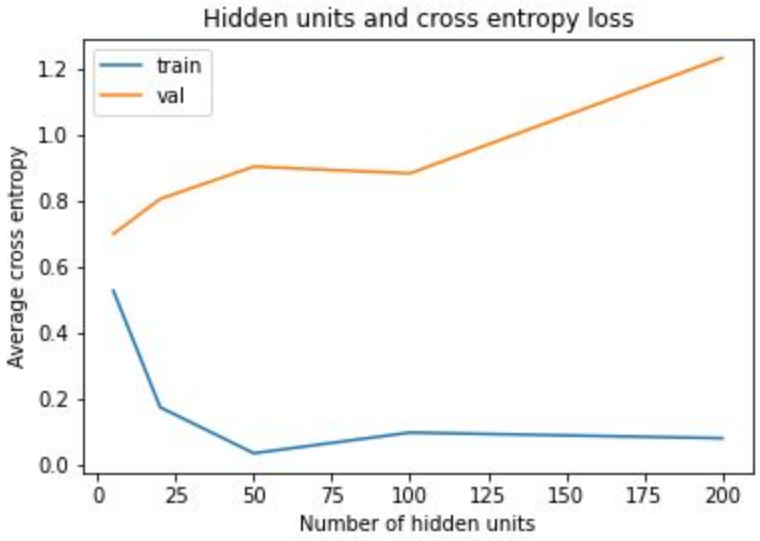
6)

Activation Function	Test Accuracy	Train Accuracy
ReLu	0.9755	1.0
sigmoid	0.9686	1.0

linear	0.9119	0.9267666666666666
tanh	0.968	0.9999833333333333

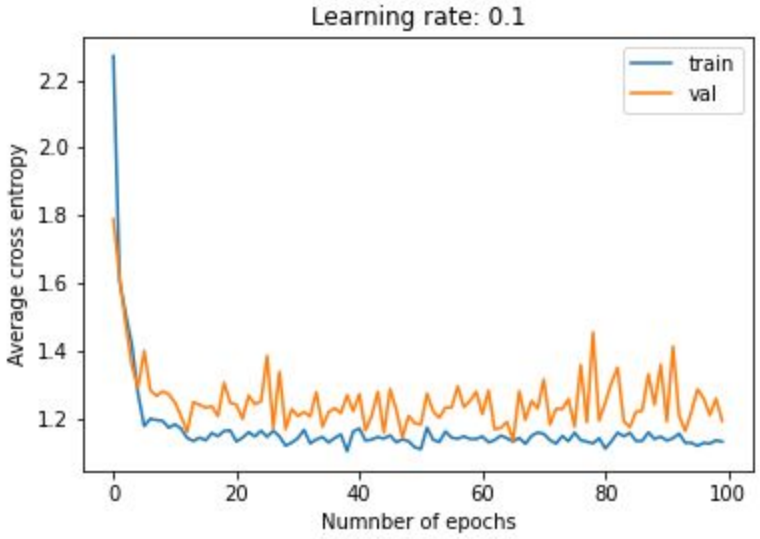
Similar accuracies are observed for my implementation and sklearn.

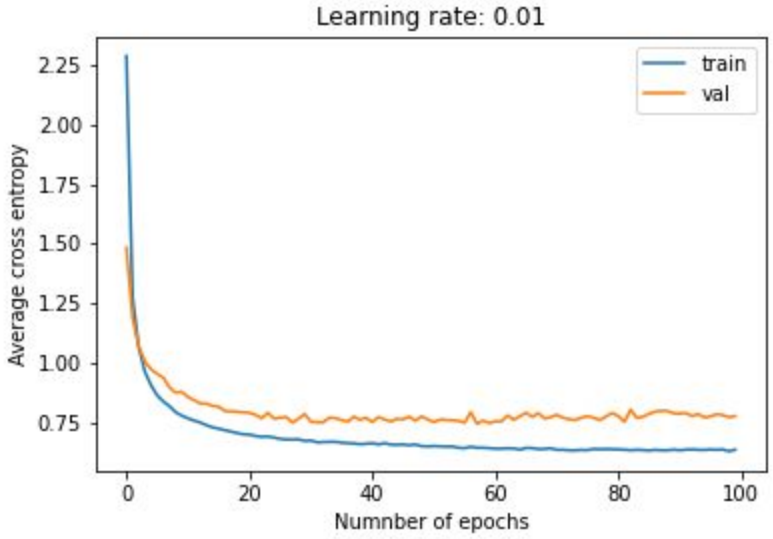
3. 1) a)

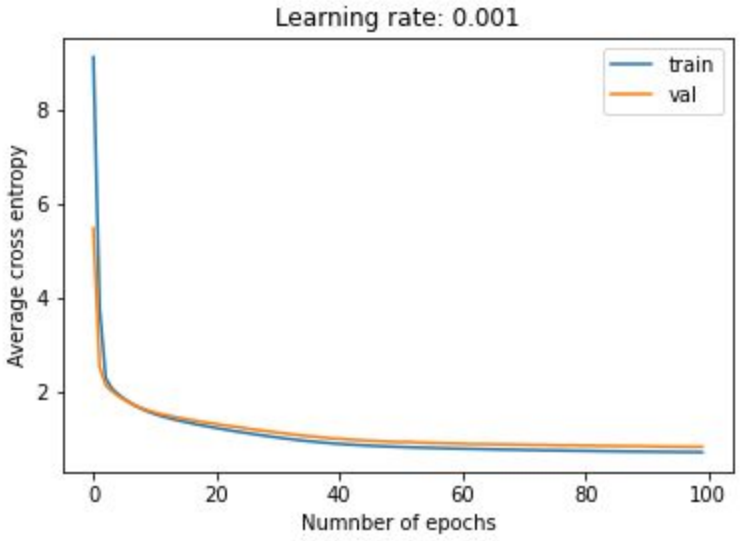


b) As the number of hidden units increases, the average training cross entropy loss decreases but the average validation cross entropy loss increases. Overfitting is happening as the number of hidden units are being

2) a)

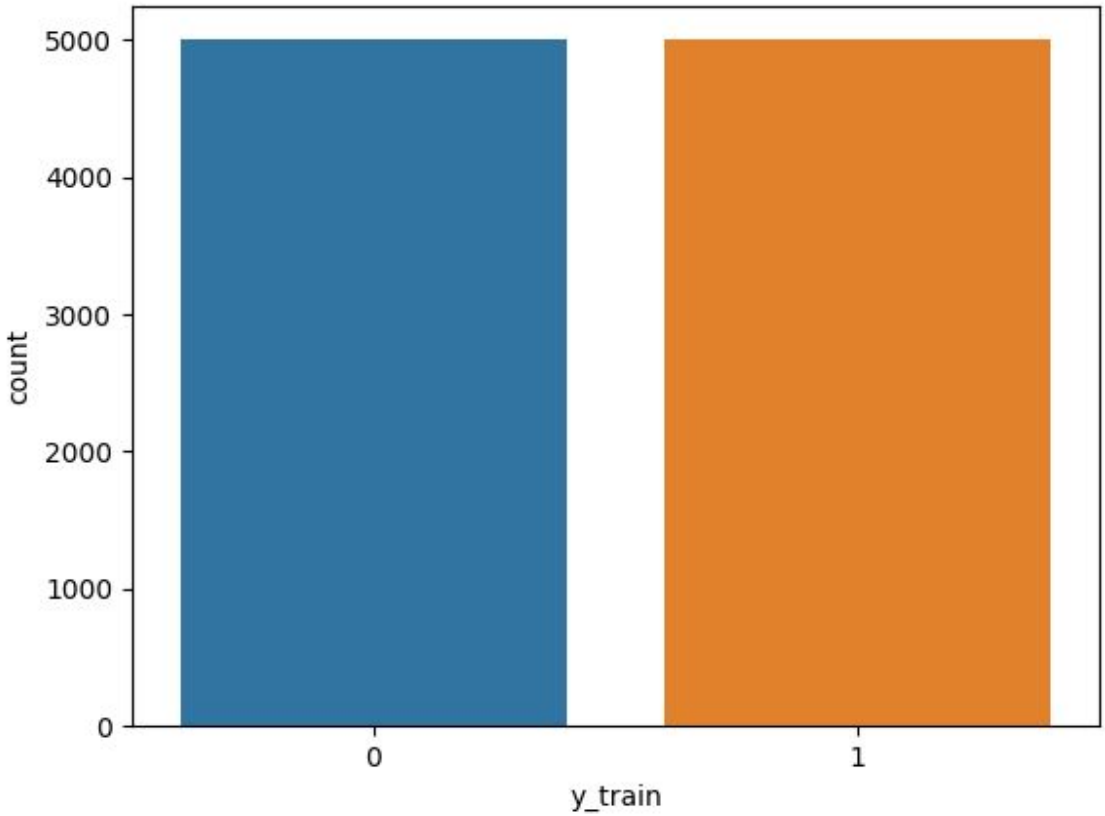

Learning rate: 0.1
train
val
Average cross entropy
2.2
2.0
1.8
1.6
1.4
1.2
0
20
40
60
80
100
Numnber of epochs

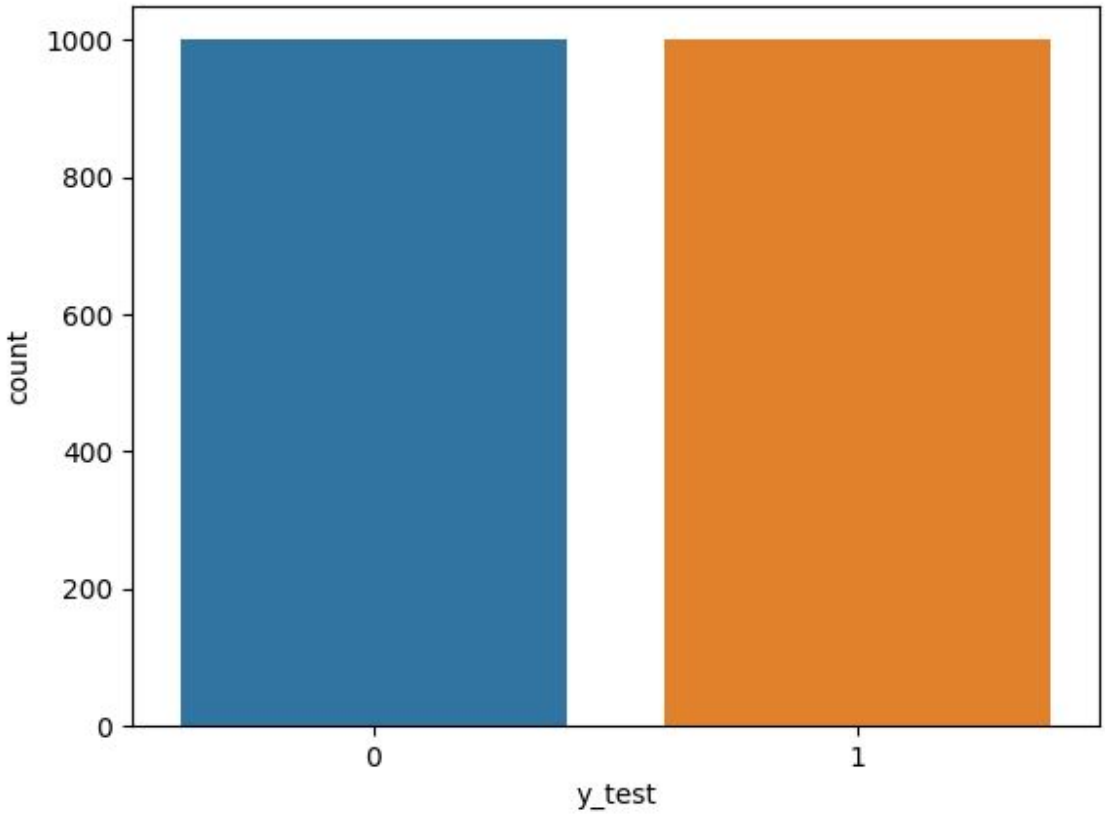

Learning rate: 0.01
train
val
Average cross entropy
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0
20
40
60
80
100
Numnber of epochs



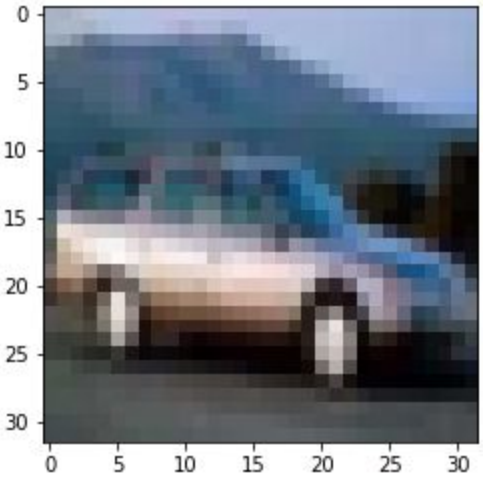
b) For learning rate 0.001, the values of average cross entropy loss are almost the same for training and validation set. The model is converging at about 60 epochs. For learning rate 0.01, average validation cross entropy loss is more than the average train cross entropy loss, but with the increase in the number of epochs it is on a whole decreasing. Convergence is reached at about 40 epochs. For learning rate 0.1, overfitting is happening. Convergence is reached at about 10 epochs. After that the plots become zigzag.

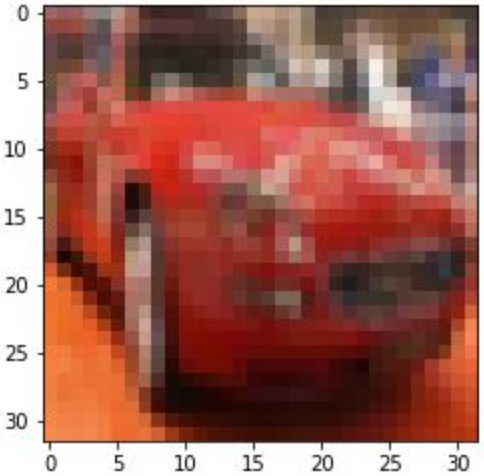
4. 1) There are 5000 instances each of classes 0 and 1 in the train data. There are 1000 instances each of classes 0 and 1 in the test data. There is an equitable distribution of data. There are 3072 feature columns.


5000
4000
3000
2000
1000
0
count
0
1
y_train



Sample images:





4) Train accuracy: 0.9818869426751592

Test accuracy: 0.95849609375

Confusion matrix:

[[965 35]

[47 953]]