

Assignment 3

Ritik Garg | 2018305

Answer1. Implemented the code

Answer2.

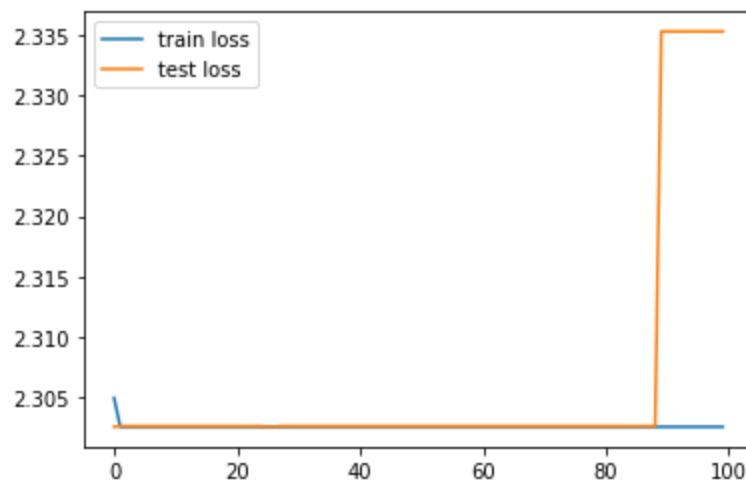
a. Weights uploaded in the drive folder:

Link:

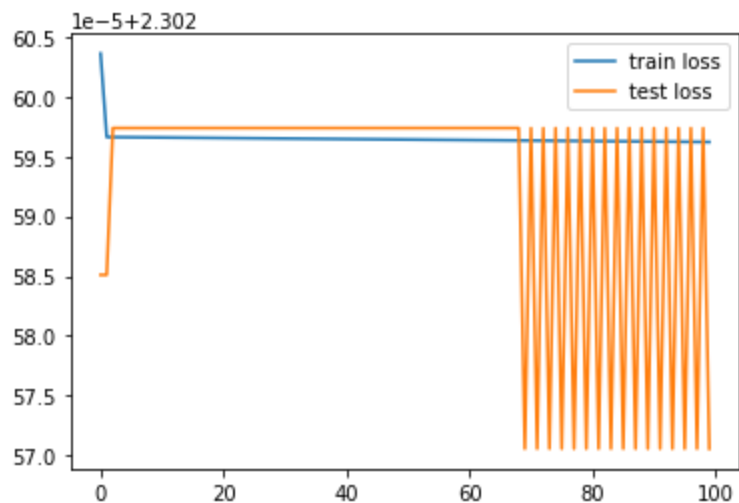
<https://drive.google.com/drive/folders/1bzphaWN3YDz3SkXhQ7f9-fTjvAHWX3d6?usp=sharing>

b. Train samples: 12000; Test samples: 2000

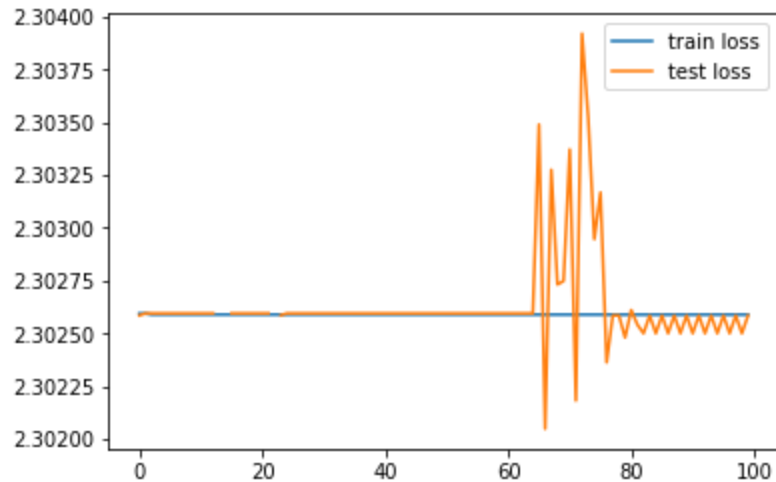
Sigmoid -> lr 0.1, bs = 50; Test score: 0.11



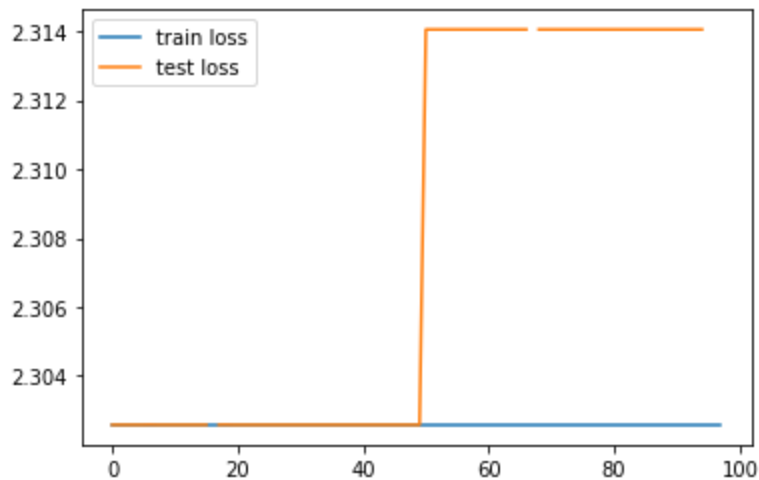
Linear: lr=0.1, bs=200; Test score: 0.159



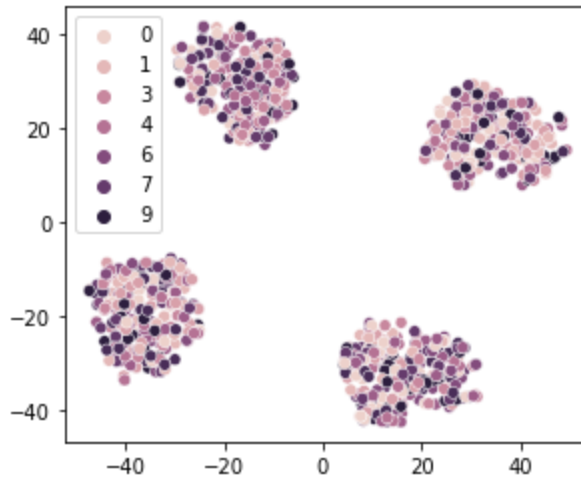
ReLU: lr=0.1, bs=200; Test score:0.1125



Tanh: lr=0.1, bs=200; Test score: 0.101



- c. The activation function of the output layer should be the softmax function. It returns the predict_proba i.e the predicted probabilities for the output layers.
- d. Total layers: 5
Hidden layers: 3
- e. The accuracy of the model is very low thats why the tsne is not able to differentiate between different classes.



f.

Relu accuracy on test data: 0.829

Sigmoid accuracy on test data: 0.918

Linear accuracy on test data: 0.886

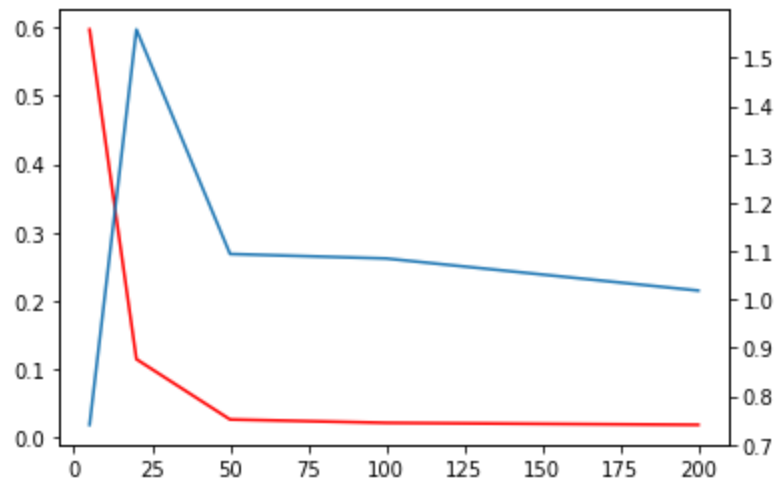
Tanh accuracy on test data: 0.918

There is some difference in the accuracy due to different ways of error handling. Sklearn provides a more optimal way for dealing with underflows and overflows.

Answer3.

1.

a.



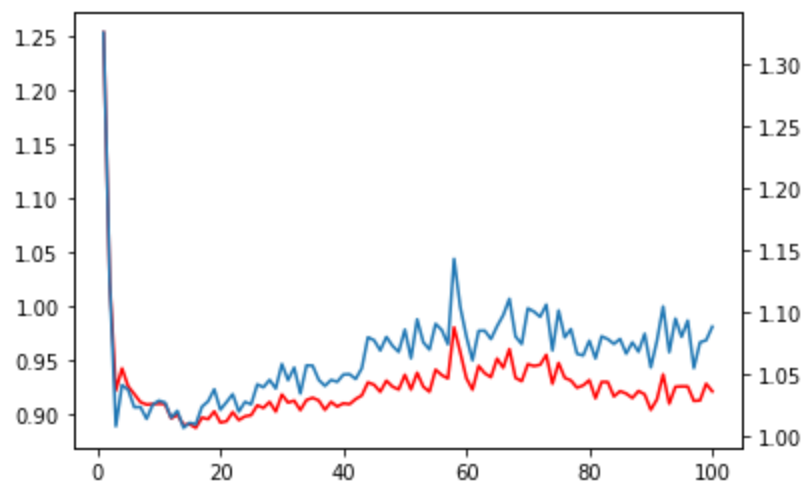
Red: training; Blue: validation

b. Reason for changes:

Changing the number of hidden layers reduces the error as seen in the plots. We minimise the loss on training data which can be seen in the graph below, but it can vary in validation loss as seen. It first increases then decreases with the number of hidden layers. Greater the number of hidden layers, more will be the learning and lesser will be the loss.

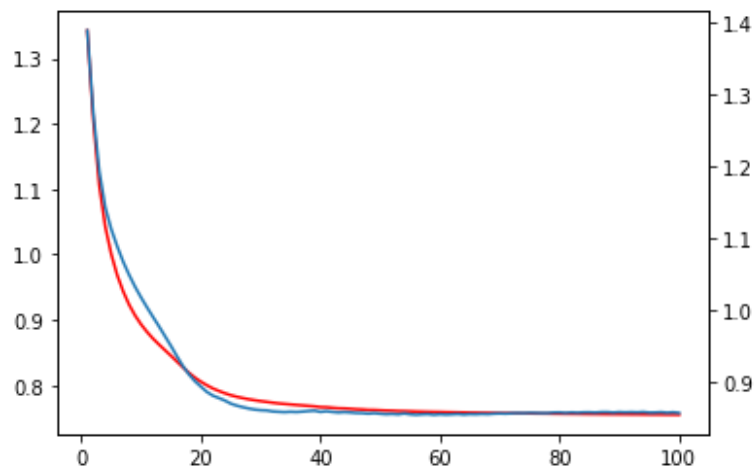
2.

a. Learning rate = 0.1



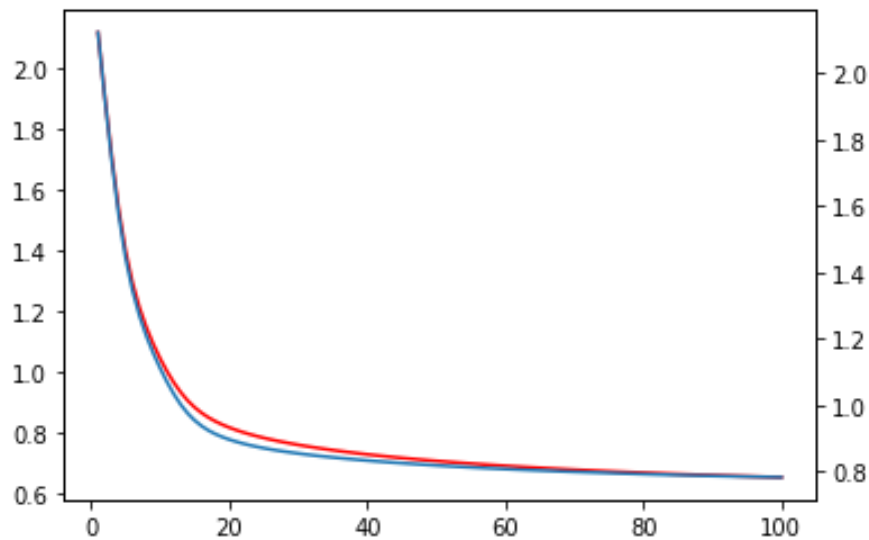
Red: training; Blue: validation

Learning rate = 0.01



Red: Training; Blue: Validation

Learning rate = 0.001



Red: Training; Blue: Validation

b. Reason for change:

We minimise the loss on training data which can be seen in the graph below, but it can vary in validation loss as seen. With the decrease in the learning rate, the graph becomes smooth and converges around $x = 20$, but it was converging early for $lr=0.1$.

Answer4.

a. EDA analysis:

(10000, 3072) shape

| | | | | | | | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|-----|
| 0 | 1 | 2 | 3 | 4 | 5 | ... | 3066 | 3067 | 3068 | 3069 | 3070 | 3071 | |
| 0 | 170 | 168 | 177 | 183 | 181 | 177 | ... | 83 | 79 | 78 | 82 | 78 | 80 |
| 1 | 159 | 150 | 153 | 154 | 138 | 184 | ... | 13 | 16 | 14 | 14 | 17 | 19 |
| 2 | 202 | 202 | 204 | 206 | 208 | 209 | ... | 212 | 227 | 237 | 243 | 243 | 243 |
| 3 | 126 | 122 | 126 | 127 | 130 | 130 | ... | 114 | 113 | 111 | 109 | 107 | 102 |
| 4 | 50 | 51 | 42 | 62 | 100 | 66 | ... | 165 | 167 | 165 | 166 | 166 | 162 |

[5 rows x 3072 columns] head of X

[1 1 0 0 1 0 1 1 1 0] head of y

RangeIndex(start=0, stop=3072, step=1) columns of X

10000 Number of samples

255 Maximum pixel value

0 Minimum pixel value

129.53166077473958 Mean pixel value

(array([0, 1]), array([5000, 5000])) Class distribution
First 9 Images



(b) and (c) part implemented in the code

(d). Accuracy: 0.9295, Auc: 0.976376

Confusion matrix: array([[927, 68],
[73, 932]])

ROC Curve

