REPORT

Assignment 3

Machine Learning - Monsoon 2020

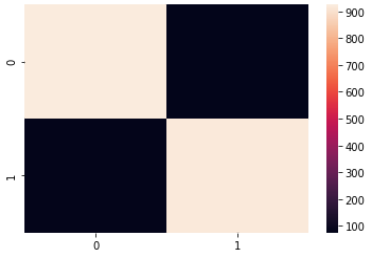
Q4.

1.

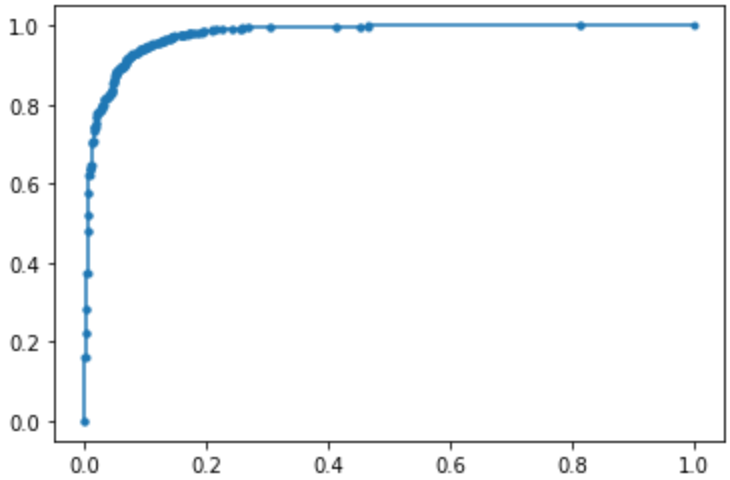
```
----Exploratory Data Analysis----
# of Samples 10000
maximum pixel value 255
minimum pixel value 0
mean pixel value 129.53166077473958
class distributions (array([0, 1]), array([5000, 5000]))
for class 0
mean pixel value 142.365784765625
max pixel value 255
min pixel value 0
for class 1
mean pixel value 116.69753678385416
max pixel value 255
min pixel value 0
first 10 labels [1 1 0 0 1 0 1 1 1 0]
---          The END          ---
```

First 9 images

1. Accuracy: 0.923
2. Confusion Matrix: `[925  75]`
   `[ 79 921]`



ROC Curve

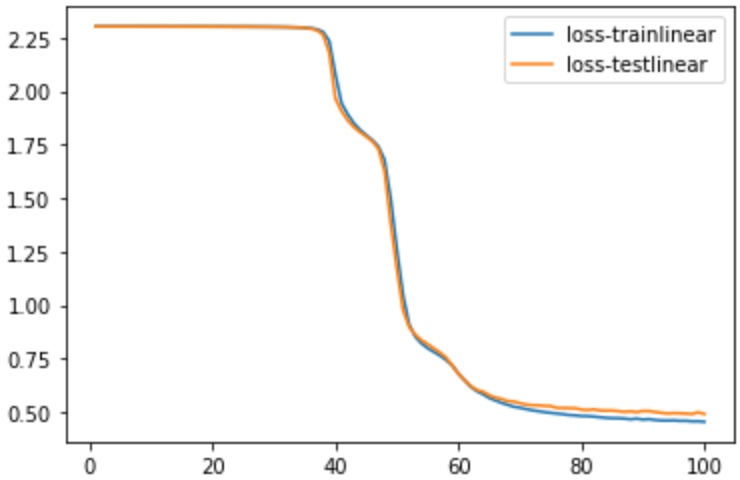


Q2:

# of training samples = 30,000

# of testing samples = 5,000

1. Part 1, 2, and 6 are combined
    1. Linear



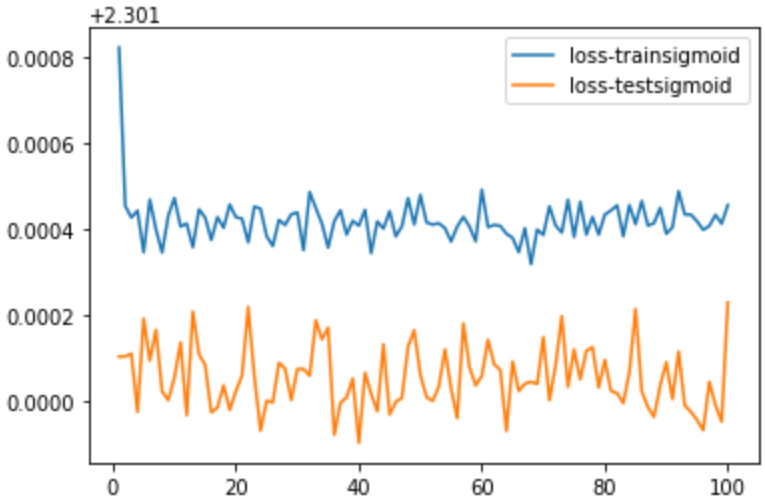
accuracy=0.865

batch_size=1500, epoch=100, LR=0.1

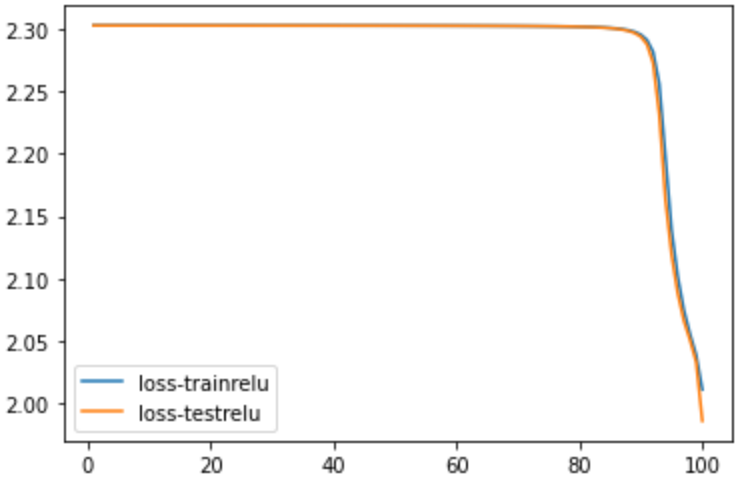
SkLearn Accuracy=0.9148

The accuracies are almost similar for both the cases since both have seemed to hit the minima. The slight difference could be accounted for by the epochs taken by both to reach the minima.

2. Sigmoid
   Accuracy = 0.1136
   Sklearn accuracy=0.1136
   Reasoning: The accuracies are exactly the same and only one class is predicted in both cases.



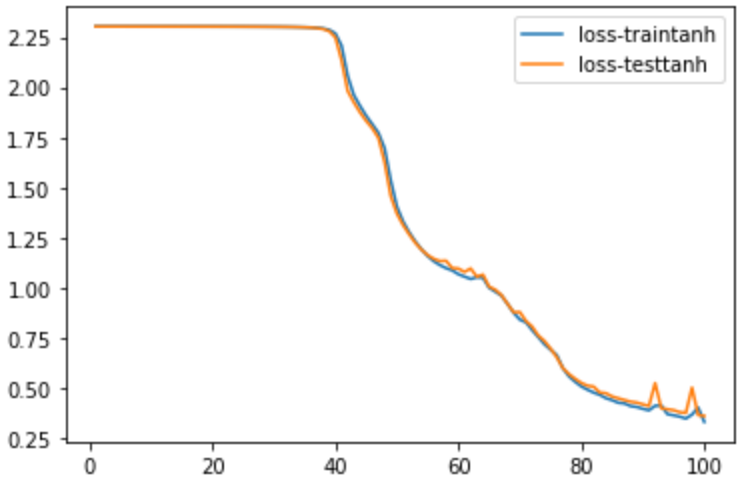
3. ReLU



   Accuracy = 0.292
   SkLearn accuracy = 0.8908
   Reasoning: there is a huge difference in the accuracy obtained from sklearn and self-implemented code. This could be accounted for by

the number of epochs taken by both the optimization algorithms(SGD in sklearn and mini-batch in self implementation). The curve is downward sloping and is yet to reach the minima.

4. Tanh



Accuracy = 0.9004
Sklearn accuracy=0.724

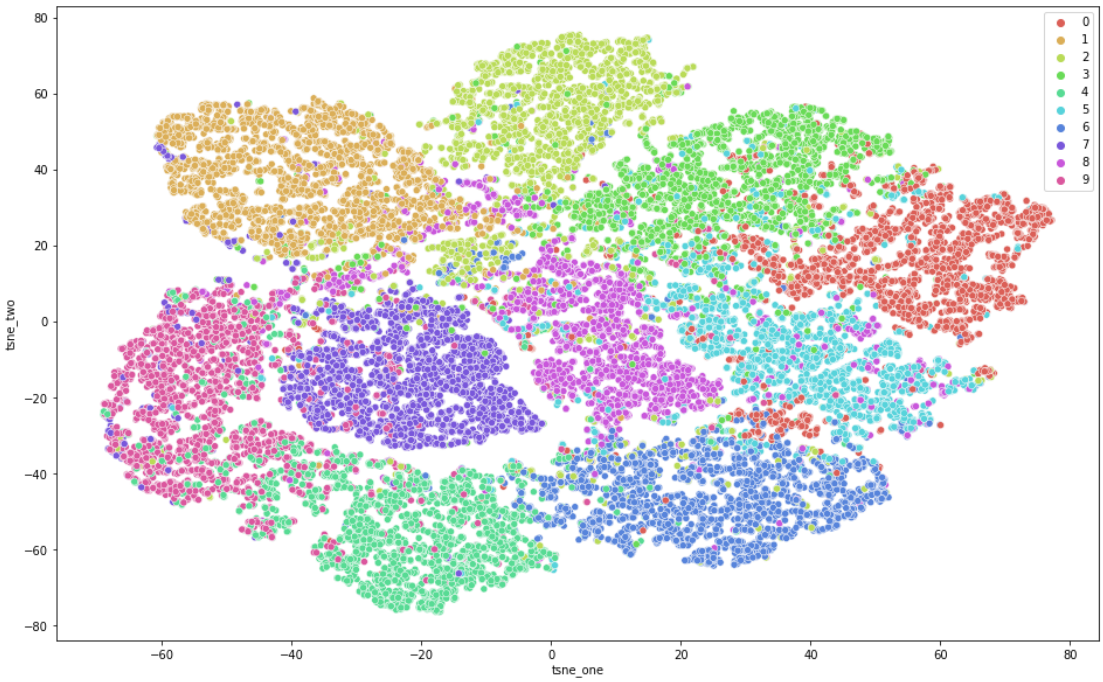
The accuracies in this case are also different because of the different optimization algorithms used in both cases. Sklearn showed a warning ‘minima has not yet reached’ with the given number of epochs.

3. The activation function that should be used in each of the cases above is the softmax function since

a. Since it gives a confidence level for each of the output classes for a given sample.
b. so that the sum of all the confidences adds up to 1 since it a multiclass problem.
c. Other activation functions like sigmoid are generally used with binary classification since it doesn’t give a probability distribution

4. Total number of layers = 5, Number of hidden layers = 3,
Input Layer = 1, Output layer = 1.

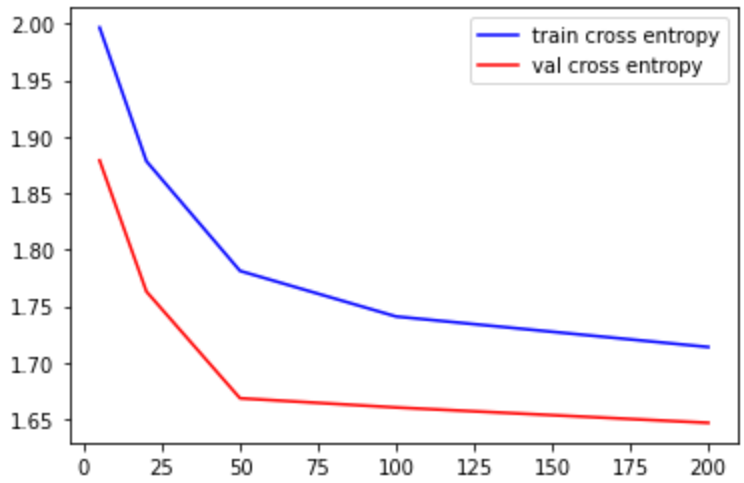
5.



Q3:

a.

Plot Hidden layers V/s number of hidden units

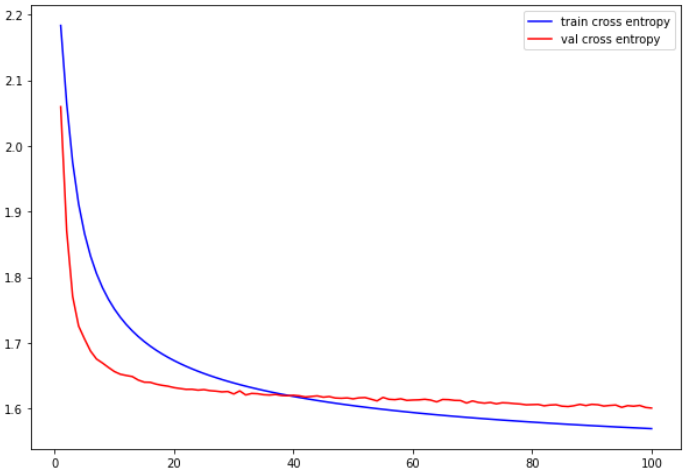


Reasoning: As the number of hidden layers increases, the complexity of the model increases. If we have **too** few **hidden units**, we will get high training error and high generalization error due to underfitting and high statistical bias. If we have **too**

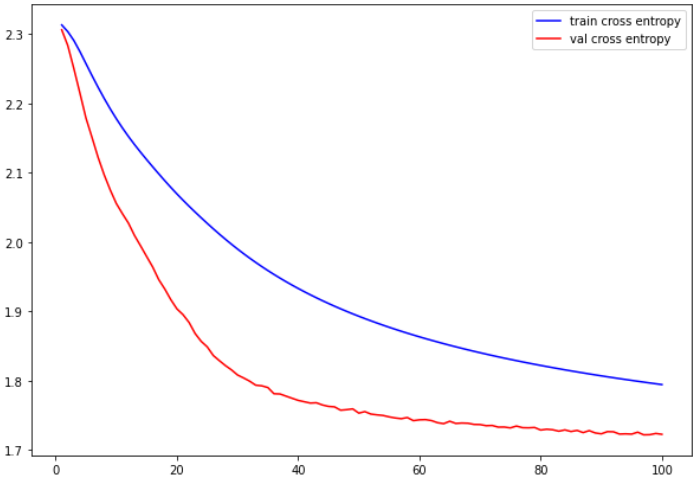
many hidden units, we get low training error but still have high generalization error due to overfitting and high variance

b.

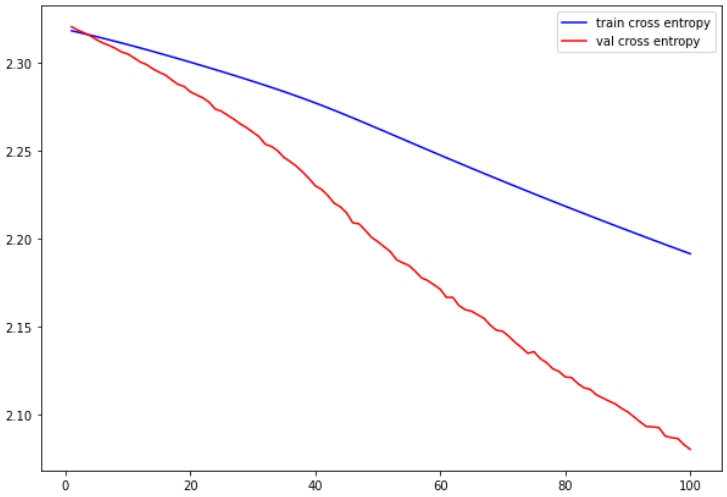
LR = 0.1



LR=0.01



LR=0.001



Reasoning: As the learning rate increases, the algorithm reaches towards the minima in a quicker fashion, while when the learning rate is low the algorithm descends towards the minima slowly and it’s pretty much visible in the plots