**SYDE 675 Assignment 3**

**Kaiyue Lu**

**21100630**

**Exercise1:**

Q2: the result of training error, training loss, test error and test loss are posted below:

A graph with a line

Description automatically generatedA graph with a line

Description automatically generatedA graph with a line

Description automatically generatedA graph with a blue line

Description automatically generated

Q3:

For linear activation, g(x) = x, the probability of number ‘4’ is calculated by:

the result is shown below:

linear activation and logistic loss

iterations: 100

training error: 0.48793

test error: 0.49297

training loss: -44848.6

test loss: -8585.1

The training error and test error are very high, and their loss are negative. Logistic regression works for probability, which means that the output of classifier should be in interval [0,1]. Thus, without the sigmoid activation function to scale output of classifier to interval [0,1], the logistic regression doesn’t work.

Q4:

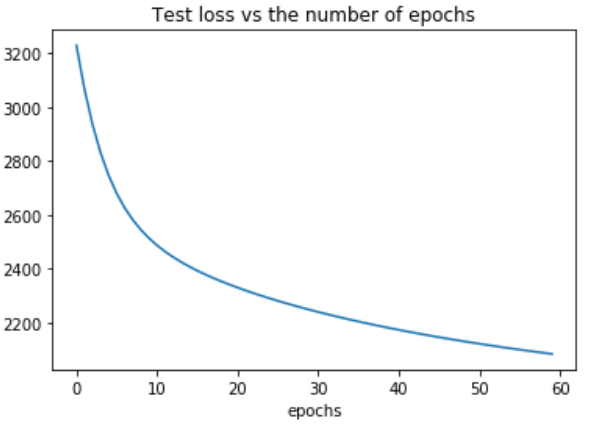
The classifier without updating bias are shown below. The result shows that their result converges but the optimal accuracy is lower than that of updating bias. The reason is that bias will increase variability of classifier to fit in datapoints. When classifier have lower flexibility, its accuracy will be worse because of its low variability.

A graph with a line

Description automatically generatedA graph with a line

Description automatically generated

A graph with a line

Description automatically generated

Q5:

The accuracy of MED classifier is 98.042%

The accuracy of MMD classifier is 98.143%

The accuracy of logistic regression is 97.2%

MMD classifier performs the best because it considers the variance of each feature of data. But if we increase the number of epochs, the accuracy of logistic regression may be higher.

**Exercise 2:**

Q2:

The result of VGG11 and SGD optimizer is shown below:

A graph of a graph of a number of epops

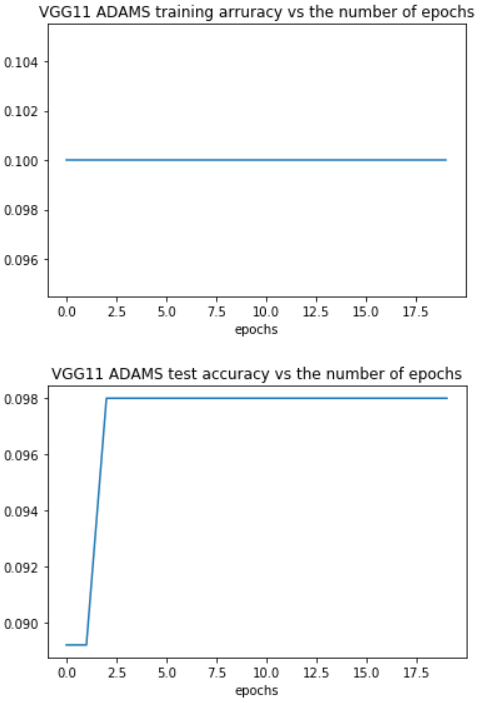
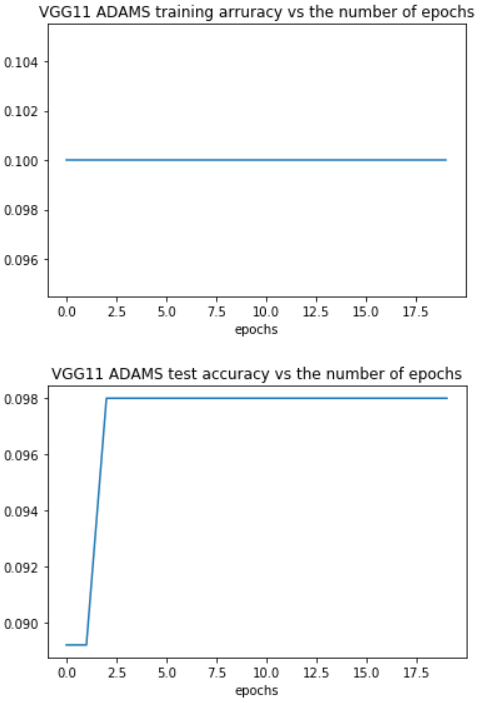
Description automatically generated with medium confidenceA graph of a number of epops

Description automatically generated

Q3:

The result of VGG11 and ADAM optimizer is shown below. The result of ADAM optimizer does not converge. There are 2 possible reasons.

* The parameter of ADAMS is not properly selected.
* The ADAM is used for preventing local minimum but will slow down the convergence at the same time.



Q4:

The result of VGG11 and Sigmoid activation is shown below. CNN with sigmoid, it performs very poor. The reason is gradient vanishing. When the input of sigmoid is very high, its gradient will be close to 0, thus the training step size is also close to 0. CNN is a kind of deep neural network, which will reinforce the effect of gradient vanishing. Compare to ReLU, the gradient of ReLU is 1 when the input is greater than 0, thus ReLU will not cause gradient vanishing.

A graph with numbers and lines

Description automatically generatedA graph with numbers and lines

Description automatically generated

Q5:

The result of VGG11 and no drop out is shown below. Without dropping out, the training process is less stable. The purpose of dropout is to decrease the overfit of classifier. With dropout the classifier will rely each of Neurons evenly. Without dropout, classifier will highly rely on certain neurons, which will cause overfit of VGG11.

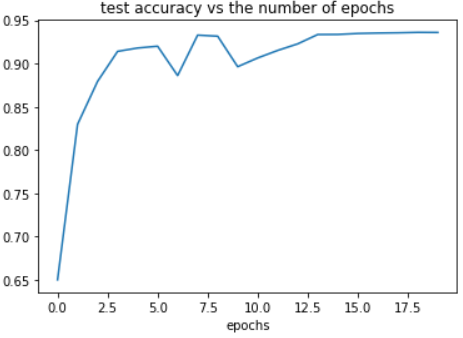
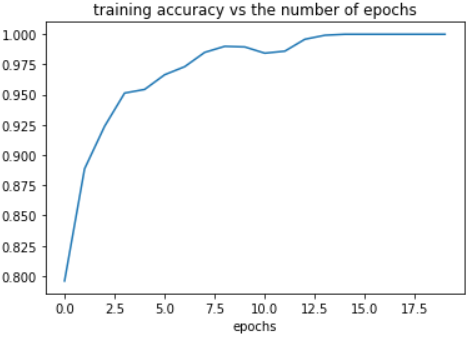
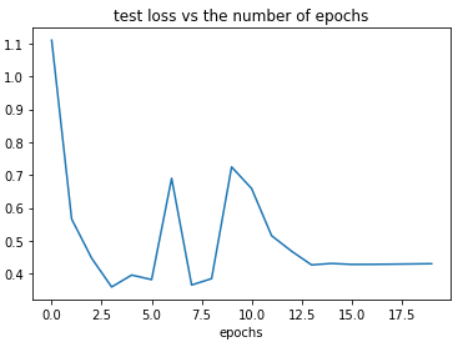
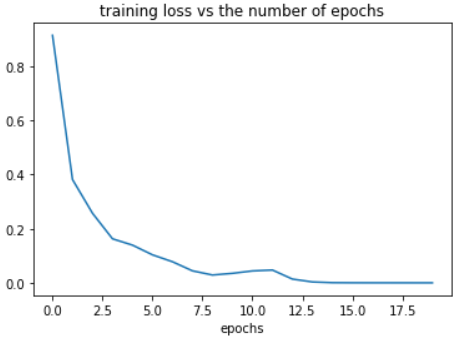
A graph of a number of epops

Description automatically generatedA graph of a number of epops

Description automatically generated

**Exercise 3:**

Q2:



Q3:

Comparing the accuracy of MLP and VGG11, VGG11 accuracy is higher. Thus, VGG11 is better. But MLP requests less parameters and it is faster for training time.

**Bouns:**

With fully connected layer, there will be