

Name: Duy Nguyen

ROB 537 – HW 1

Using training set 1

A. Hidden Unit and Time training

I train the data in 500 epochs with the hidden unit goes from 0 to 90 with a 10 units increment. Learning rate is at 0.0002, momentum = 0.7, L2 penalty = 0.0002

As the training time increase, the accuracy also increases. However, when the training process reaches a maximal point, the accuracy stops increasing and becomes stables. For the hidden units, from 0 to around 50 hidden units, the rate of going to the stable accuracy increases. However, after 50 hidden units, the rate of going to the stable point starts to decrease. It happens to all 3-test data. However, looking at the result when there is no hidden unit (the program picks the same value for all the answer), I can see that the test 2 and test 3 are extremely bias and in different direction.

B. Learning Rate

I train the data in 500 epochs with the learning rate goes from 0 to 0.0018 with a 0.0002 increment. Hidden unit is 50, momentum = 0.7, L2 penalty = 0.0002

As the learning rate increases, the converging tends to increase. However, the learning rate does not change that much after 0.0006. I guess that my learning rate increment is a little small to see any significant changes in different learning rate. Test 2 and test 3 fluctuate more around the stable accuracy (85%) before converging to its than test 1. I guess that is because test 2 and 3 are more bias than test 1

C. Momentum

I train the data in 500 epochs with the momentum goes from 0 to 0.9 with a 0.1 increment. Hidden unit is 50, learning rate = 0.0008, L2 penalty = 0.0002

In general, I can see that the converging rate increases as the momentum increases. However, that increment becomes insignificant after 0.6. Test 1 tends to converge faster than test 2 and test 3. Test 2 and test 3 fluctuate around the converging point more than test 1. Again, the reason could be because of the bias of test 2 and test 3

D. L2 Penalty

I train the data in 500 epochs with the L2 penalty goes from 0 to 0.0018 with a 0.0002 increment. Hidden unit is 50, learning rate = 0.0008, momentum = 0.7

I think there is an increment in the converging rate as L2 Penalty increase. However, that increment is not that much. Test 1 tends to converge faster than test 2 and test 3.

Using training set 2

A. Hidden Unit and Time training

I train the data in 500 epochs with the hidden unit goes from 0 to 90 with a 10 units increment. Learning rate is at 0.0002, momentum = 0.7, L2 penalty = 0.0002

As the training time increase, the accuracy also increases. However, in test 1 and test 3 the algorithm takes a lot longer to learn than training set 1. In the contract, test 2 takes almost no time to learn the correct function. The rate of going to the stable point increase around 40 hidden units for test 1 and 10 hidden units for test 3. The rate of going to the stable points almost the same for all different hidden units in test 2.

B. Learning Rate

I train the data in 500 epochs with the learning rate goes from 0 to 0.0018 with a 0.0002 increment. Hidden unit is 50, momentum = 0.7, L2 penalty = 0.0002

For test 1, I can see clearly that as the learning rate increases, the rate of converging increases. The converging point is around 65 % accuracy. For test 2, there is no effect of the learning rate on the data, it starts off with the really high accuracy and the stable point is around 91% accuracy. For test 3, the learning rate have similar influence on the converging rate as test 1, however, the converging accuracy is around 40%. I think is because of the bias in training and test data sets

C. Momentum

I train the data in 500 epochs with the momentum goes from 0 to 0.9 with a 0.1 increment. Hidden unit is 50, learning rate = 0.0008. L2 penalty = 0.0002

There is a clear increment in the converging rate of test 1 and test 3 when the momentum increases. There is no effect of momentum on the converging for test 2. Test 2 has the highest stable accuracy (around 91%), test 1 is the second (around 65 %), and test 3 has the lowest stable accuracy (around 40). Again, I think it is due to the bias of the training and test data. I think is because of the bias in training and test data sets

D. L2 Penalty

I train the data in 500 epochs with the L2 penalty goes from 0 to 0.0018 with a 0.0002 increment. Hidden unit is 50, learning rate = 0.0008, momentum = 0.7

L2 Penalty has a lot influences on test 1 and test 3 result as there is big improvement in the converging rate as L2 increases. L2 has no influence on the converging rate of test 2. Test 2 has the highest stable accuracy (around 91%), test 1 is the second (around 65 %), and test 3 has the lowest stable accuracy (around 40). Again, I think it is due to the bias of the training and test data. I think is because of the bias in training and test data sets

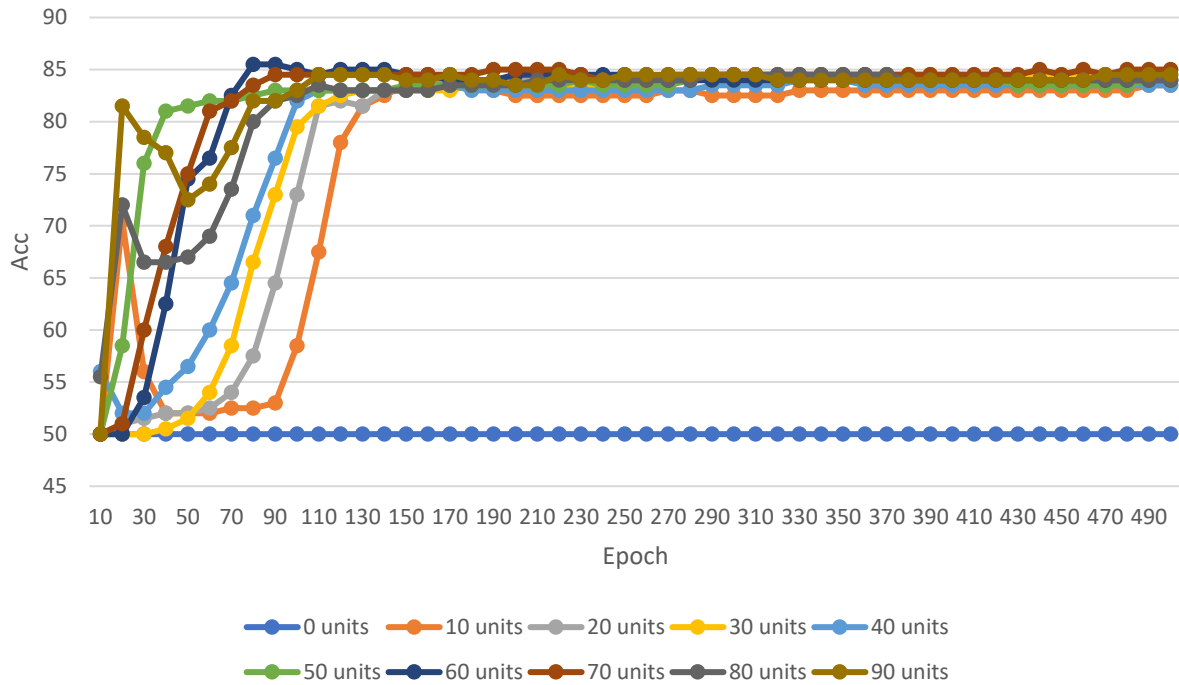
Conclusion

I can conclude that the training set 2 is bias and its bias is similar to the test 2 because it takes extremely long time to train in a balanced data set and extremely fast in a similar bias data. When the training data is not bias, there is not a lot influences from Momentum and L2 Penalty, however, when the training data is bias, Momentum and L2 Penalty improve the performance of the algorithm a lot.

I think that one way to fit the problem is to combine the two-training data set together and flip the other of them every time we run a new epoch. Additionally, we can select small batch of that combined data by randomly pick with replacement some element in the training set and do the training with its.

NOTE: all the graph data is next page

Epoch vs Accuracy of different Hidden Unit Value in test 1 using train set 1



Epoch vs Accuracy of different Hidden Unit Value in test 2 using train set 1

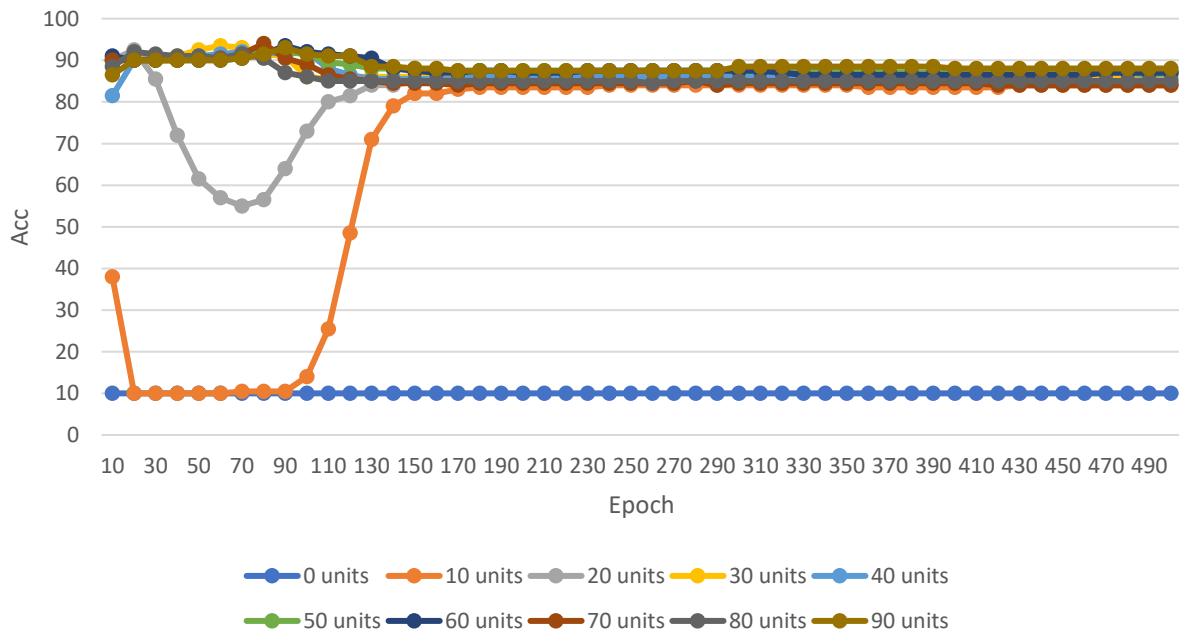


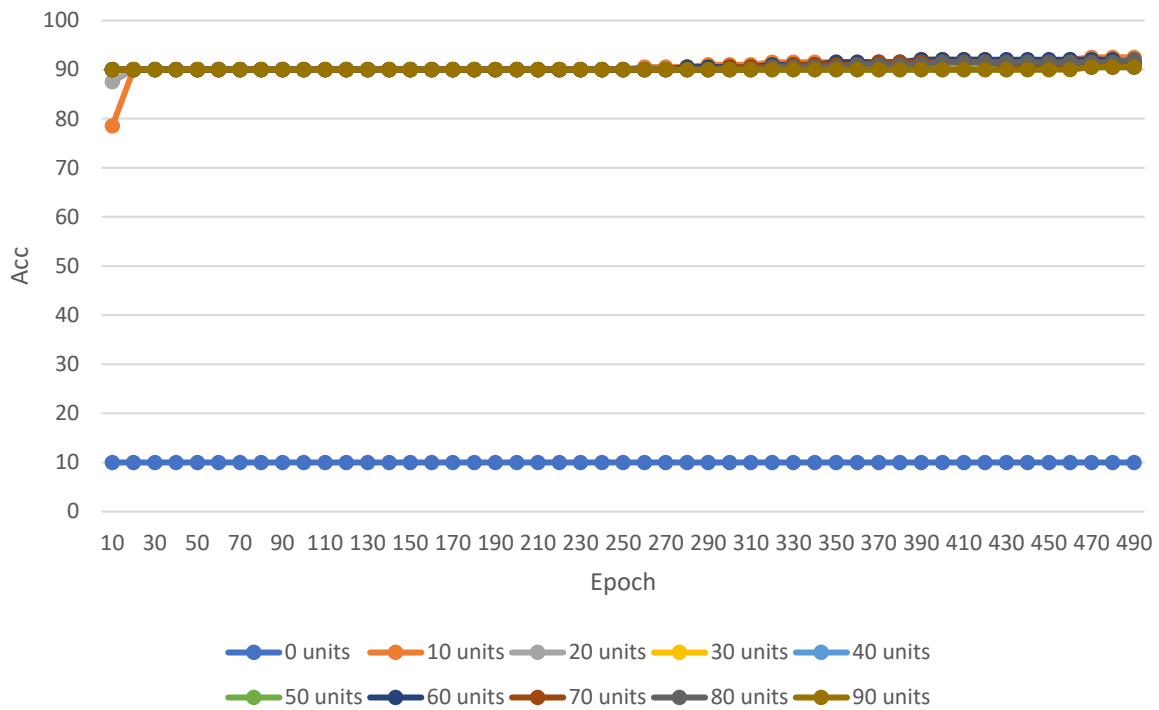
Figure 1 is a line graph showing Accuracy (Acc) on the y-axis (ranging from 0 to 100) versus Epoch on the x-axis (ranging from 10 to 490). The graph displays the performance of a model with different numbers of units (0, 10, 20, 30, 40, 50, 60, 70, 80, 90) over time. The legend indicates the following series:

- 0 units (Blue line with circles)
- 10 units (Orange line with circles)
- 20 units (Grey line with circles)
- 30 units (Yellow line with circles)
- 40 units (Light blue line with circles)
- 50 units (Green line with circles)
- 60 units (Dark blue line with circles)
- 70 units (Brown line with circles)
- 80 units (Dark grey line with circles)
- 90 units (Olive line with circles)

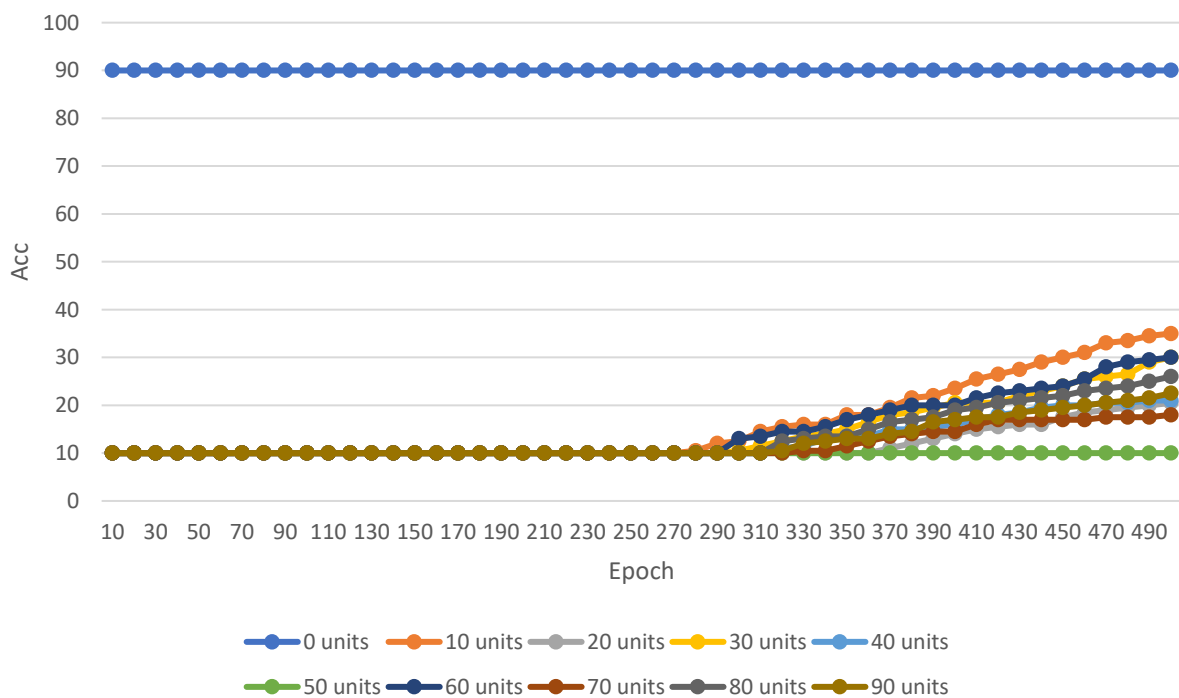
The graph shows that accuracy generally increases with the number of units. The 0 units series (blue) reaches the highest accuracy, stabilizing around 90%. The 10 units series (orange) reaches the lowest accuracy, stabilizing around 78%. The 20 units series (grey) and 30 units series (yellow) reach similar accuracy levels around 80%. The 40 units series (light blue) reaches an accuracy of about 78% by epoch 150. The 50 units series (green) reaches an accuracy of about 80% by epoch 150. The 60 units series (dark blue) reaches an accuracy of about 80% by epoch 150. The 70 units series (brown) reaches an accuracy of about 80% by epoch 150. The 80 units series (dark grey) reaches an accuracy of about 80% by epoch 150. The 90 units series (olive) reaches an accuracy of about 80% by epoch 150.

Epoch	0 units	10 units	20 units	30 units	40 units	50 units	60 units	70 units	80 units	90 units
10	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
100	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
200	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
300	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
400	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
500	50.0	58.5	61.5	61.5	67.0	57.0	60.0	64.0	64.0	65.0

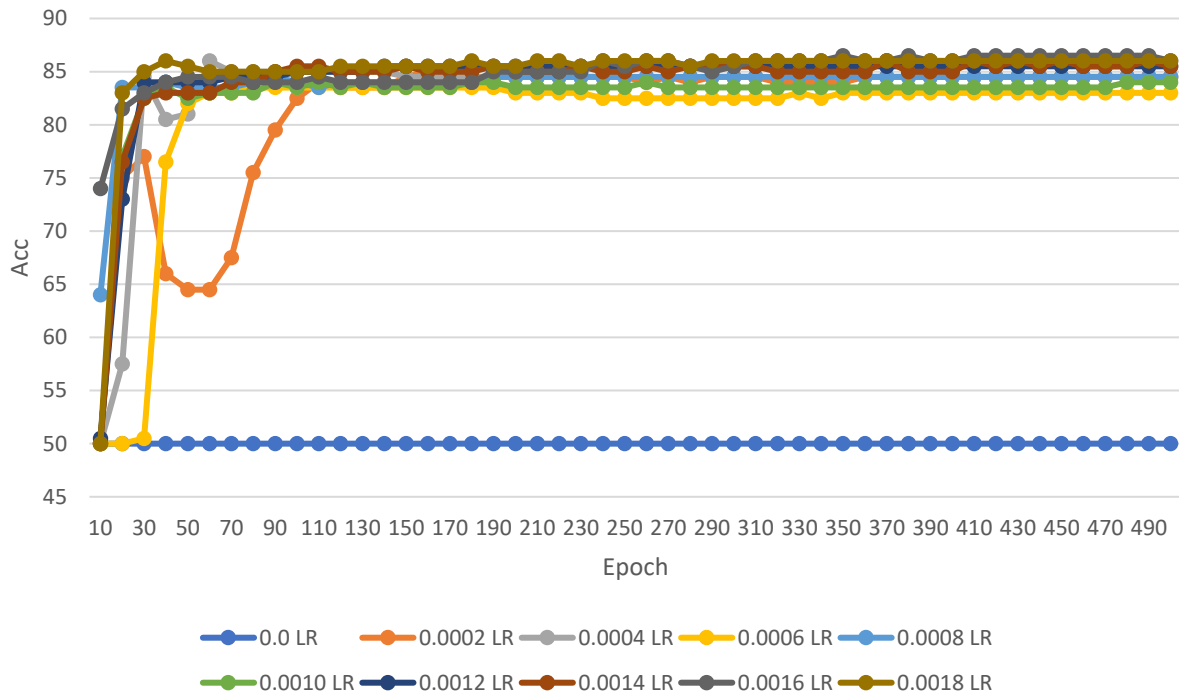
Epoch vs Accuracy of different Hidden Unit Value in test 2
using train set 2



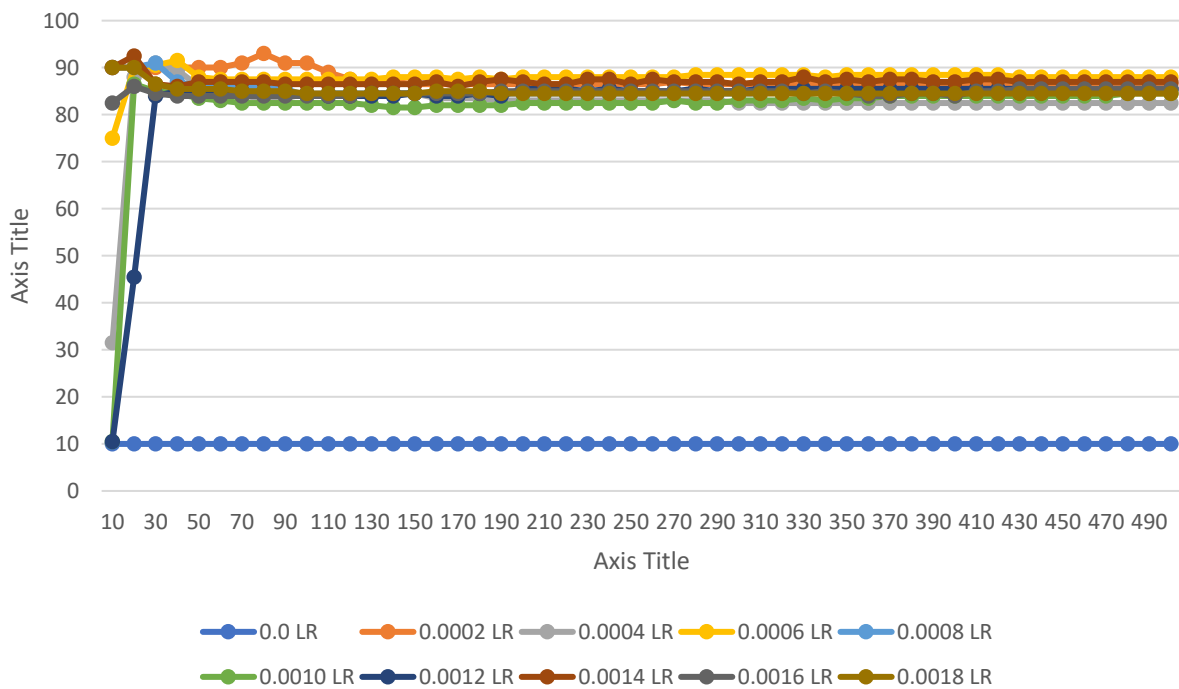
Epoch vs Accuracy of different Hidden Unit Value in test 3 using
train set 2



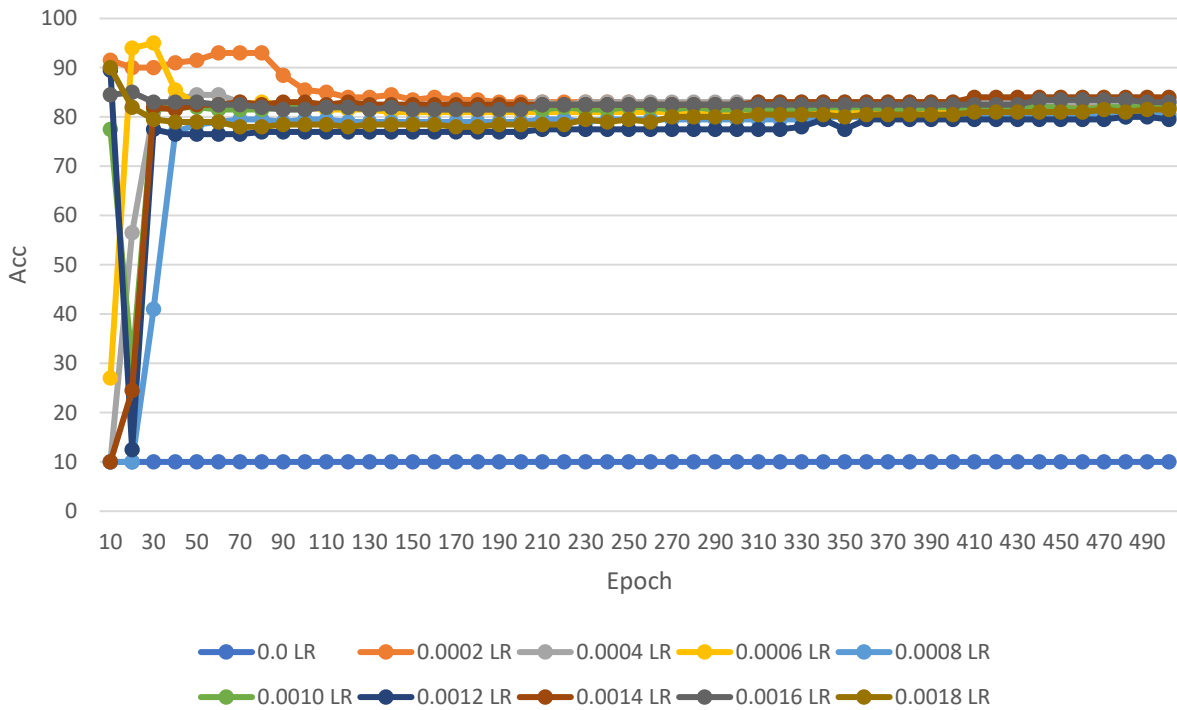
Epoch vs Accuracy of different Learning Rate Value in test 1 using train set 1



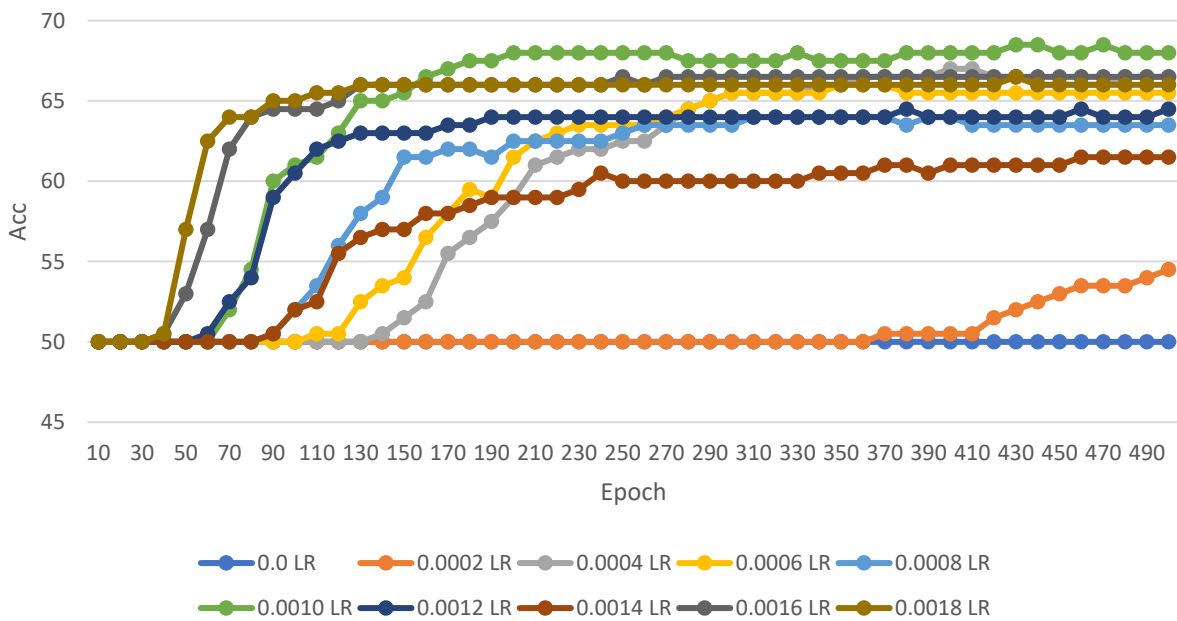
Epoch vs Accuracy of different Learning Rate Value in test 2 using train set 1

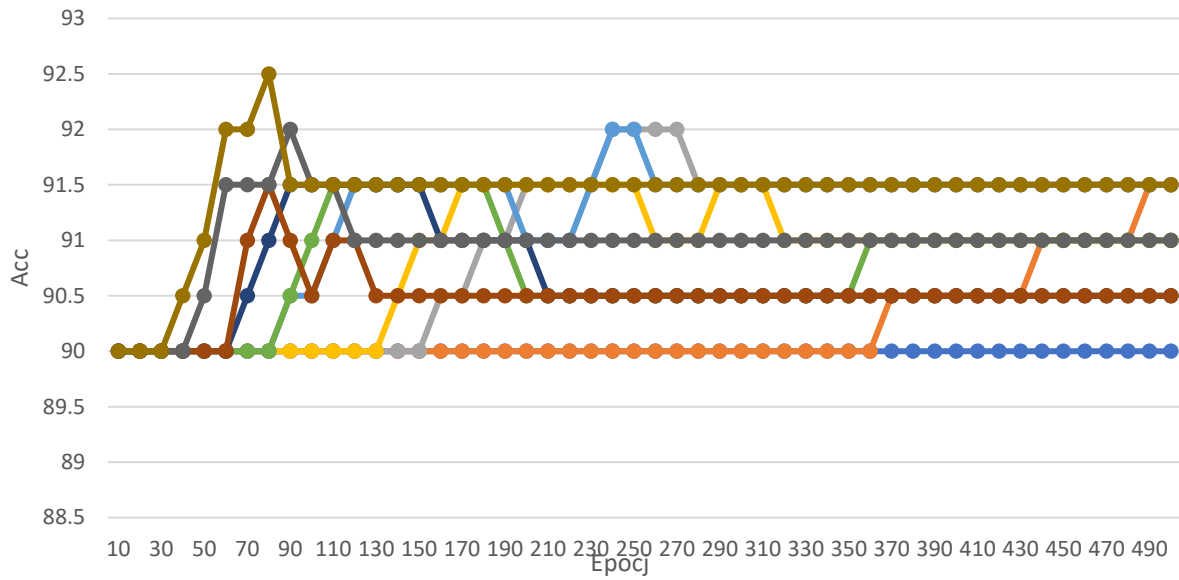


Epoch vs Accuracy of different Learning Rate Value in test 3 using train set 1



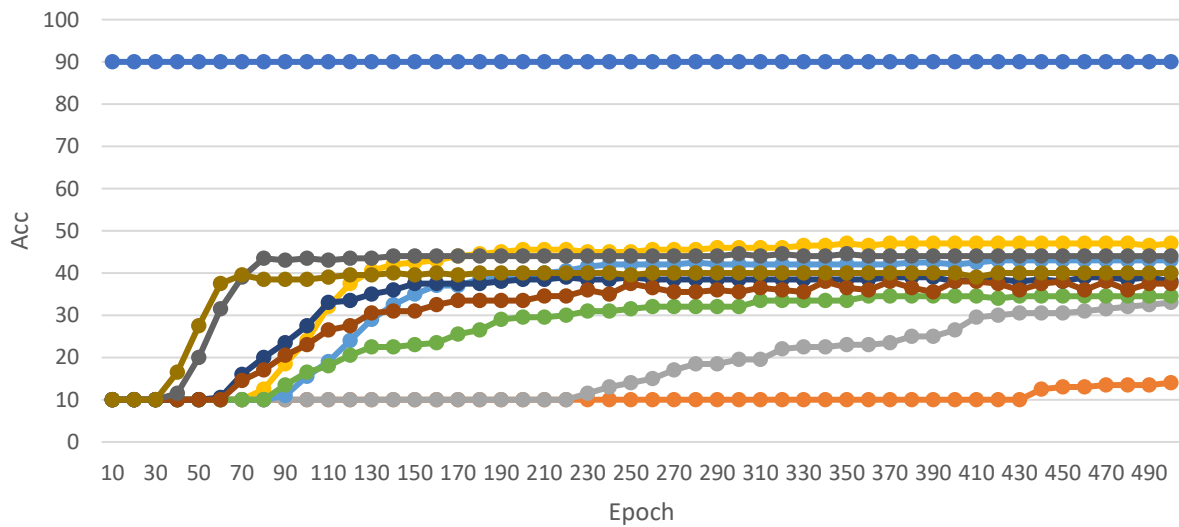
Epoch vs Accuracy of different Learning Rate Value in test 1 using train set 2



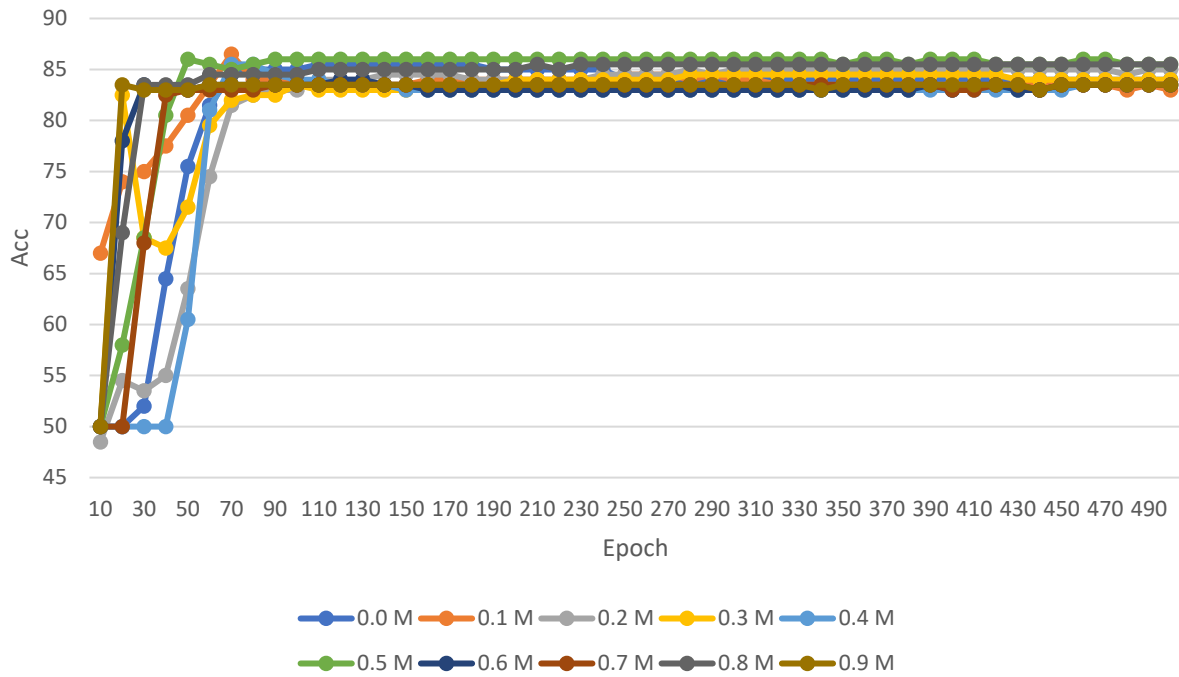


The graph displays the training accuracy of several models over 490 epochs. The blue line at the top represents a baseline or a highly accurate model that remains constant. The other models show a rapid increase in accuracy in the first 100-150 epochs, followed by a plateau. The yellow and grey lines achieve the highest accuracy among the group, while the orange line shows the lowest performance.

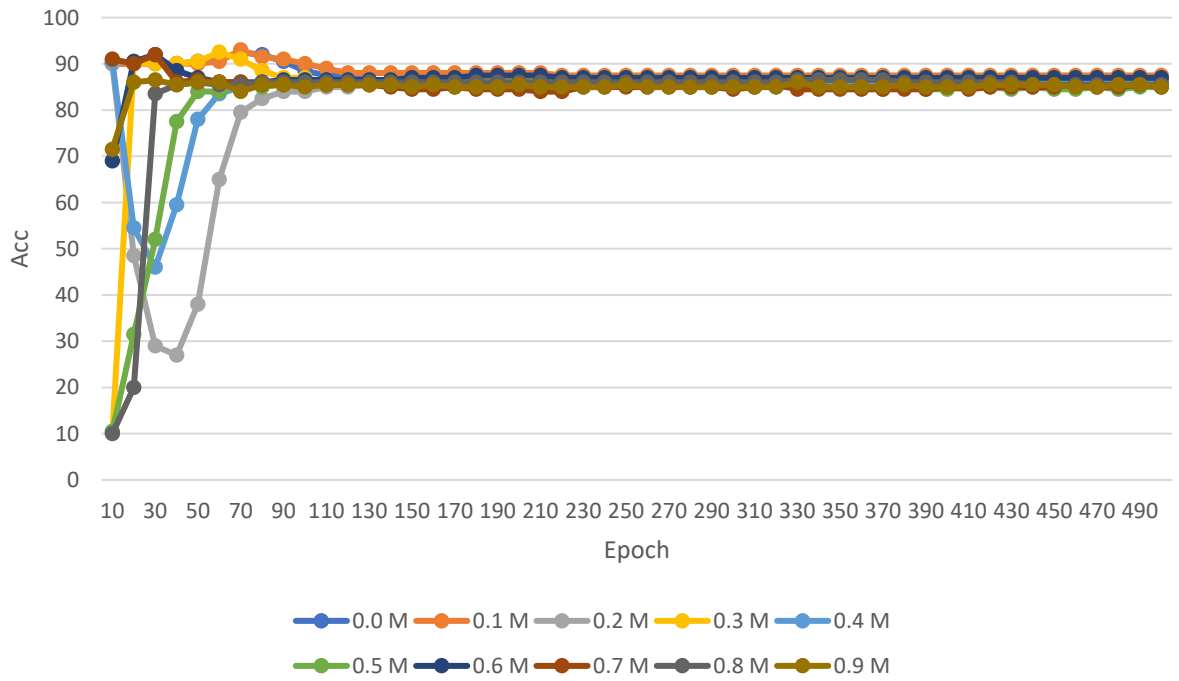
Epoch	Blue (Top)	Yellow	Grey	Dark Blue	Light Blue	Brown	Green	Orange (Bottom)
10	90	10	10	10	10	10	10	10
50	90	38	30	15	10	10	10	10
100	90	40	43	25	15	15	10	10
150	90	42	44	35	30	25	20	10
200	90	45	45	38	35	30	25	10
250	90	45	45	38	35	30	25	10
300	90	45	45	38	35	30	25	10
350	90	45	45	38	35	30	25	10
400	90	45	45	38	35	30	25	10
450	90	45	45	38	35	30	25	10
490	90	45	45	38	35	30	25	10



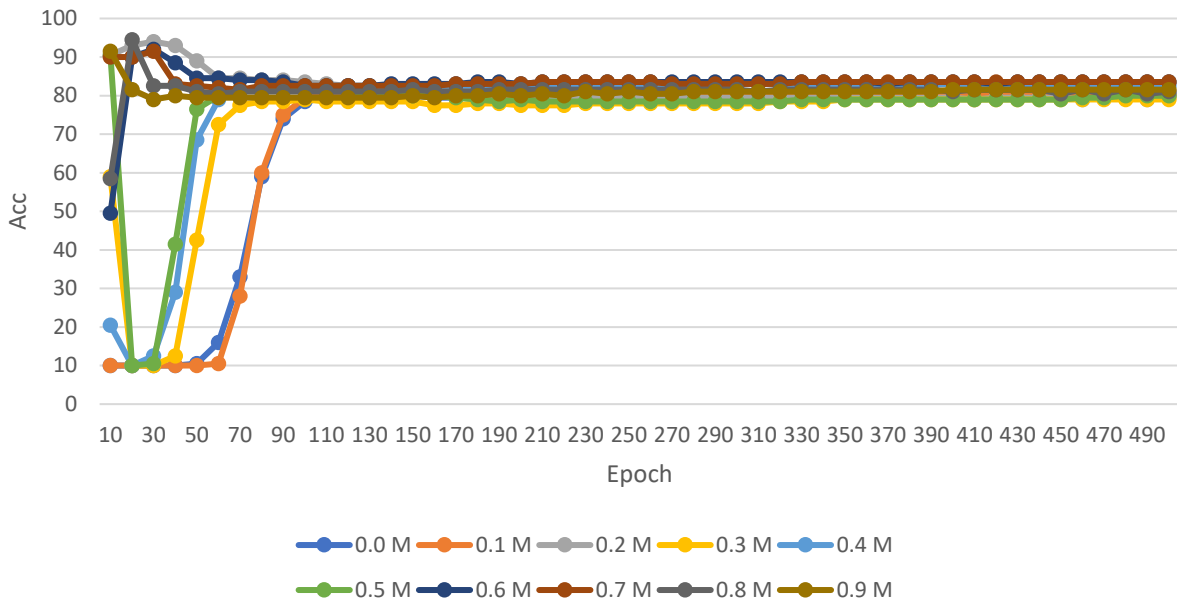
Epoch vs Accuracy of different Momentum Value in test 1 using train set 1



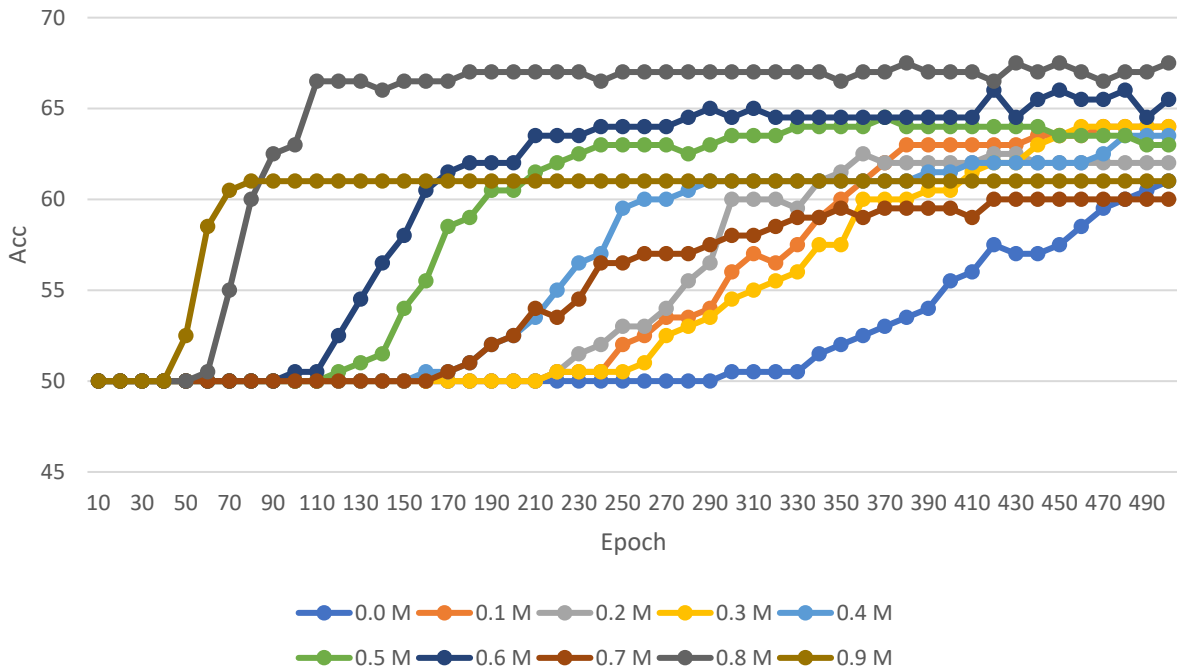
Epoch vs Accuracy of different Momentum Value in test 2 using train set 1



Epoch vs Accuracy of different Momentum Value in test 3 using train set 1



Epoch vs Accuracy of different Momentum Value in test 1 using train set 2

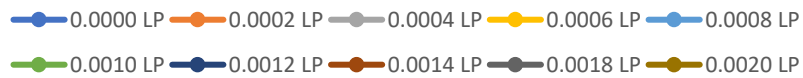


[illegible]

Figure 1 is a line plot showing the training accuracy (Acc) of ViT models over 490 epochs for different memory budgets (0.0 M to 0.9 M). The y-axis represents Accuracy (Acc) from 0 to 50, and the x-axis represents Epochs from 10 to 490. The legend indicates the memory budget for each line: 0.0 M (blue), 0.1 M (orange), 0.2 M (grey), 0.3 M (yellow), 0.4 M (light blue), 0.5 M (green), 0.6 M (dark blue), 0.7 M (brown), 0.8 M (dark grey), and 0.9 M (olive). The plot shows that as the memory budget increases, the final accuracy also increases. Models with 0.8 M and 0.9 M memory reach the highest accuracy of approximately 45-46%, while models with 0.0 M and 0.1 M memory reach the lowest accuracy of approximately 30-32%.

The graph displays the accuracy of several models over 490 epochs. The y-axis represents Accuracy (Acc) from 45 to 90, and the x-axis represents Epochs from 10 to 490. The models, represented by different colored lines, all show a rapid increase in accuracy from epoch 10 to 50, followed by a plateau. The models are color-coded: blue, orange, green, red, yellow, and grey.

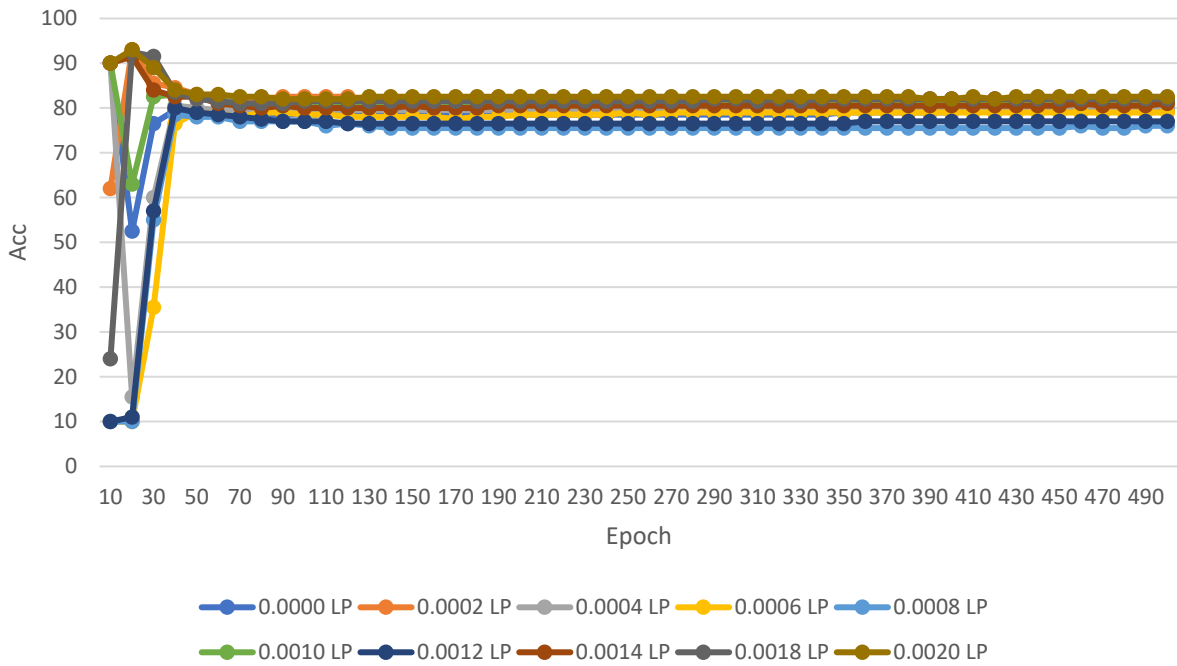
Epoch	Blue	Orange	Green	Red	Yellow	Grey
10	50	50	50	50	50	50
20	50	50	50	50	50	50
30	50	50	50	50	50	50
40	50	50	50	50	50	50
50	50	50	50	50	50	50
60	50	50	50	50	50	50
70	50	50	50	50	50	50
80	50	50	50	50	50	50
90	50	50	50	50	50	50
100	50	50	50	50	50	50
110	50	50	50	50	50	50
120	50	50	50	50	50	50
130	50	50	50	50	50	50
140	50	50	50	50	50	50
150	50	50	50	50	50	50
160	50	50	50	50	50	50
170	50	50	50	50	50	50
180	50	50	50	50	50	50
190	50	50	50	50	50	50
200	50	50	50	50	50	50
210	50	50	50	50	50	50
220	50	50	50	50	50	50
230	50	50	50	50	50	50
240	50	50	50	50	50	50
250	50	50	50	50	50	50
260	50	50	50	50	50	50
270	50	50	50	50	50	50
280	50	50	50	50	50	50
290	50	50	50	50	50	50
300	50	50	50	50	50	50
310	50	50	50	50	50	50
320	50	50	50	50	50	50
330	50	50	50	50	50	50
340	50	50	50	50	50	50
350	50	50	50	50	50	50
360	50	50	50	50	50	50
370	50	50	50	50	50	50
380	50	50	50	50	50	50
390	50	50	50	50	50	50
400	50	50	50	50	50	50
410	50	50	50	50	50	50
420	50	50	50	50	50	50
430	50	50	50	50	50	50
440	50	50	50	50	50	50
450	50	50	50	50	50	50
460	50	50	50	50	50	50
470	50	50	50	50	50	50
480	50	50	50	50	50	50
490	50	50	50	50	50	50



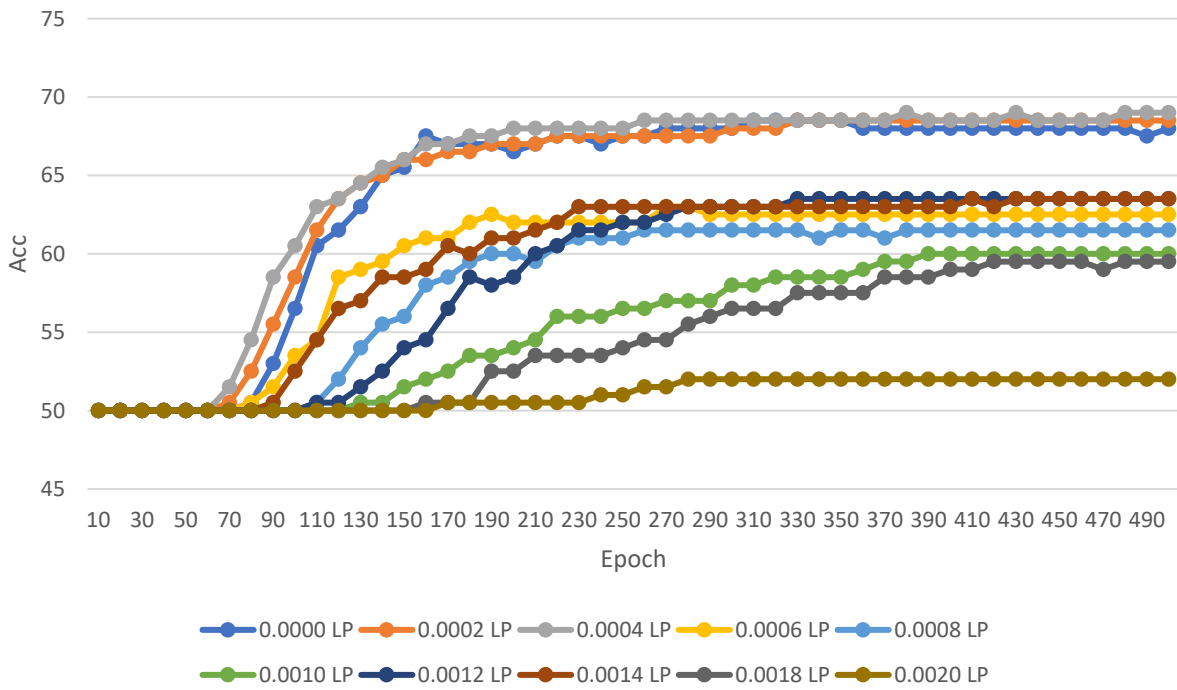
The graph displays the training accuracy (Acc) of various models over 490 epochs. The y-axis represents accuracy from 0 to 100, and the x-axis represents the epoch number. Most models show a sharp increase in accuracy within the first 30 epochs, followed by a plateau. The models are color-coded: a green line for 100% accuracy, a blue line for 90%, an orange line for 85%, a brown line for 80%, a yellow line for 75%, and a grey line for the baseline. The baseline model shows a much slower increase in accuracy, reaching approximately 85% by epoch 490.



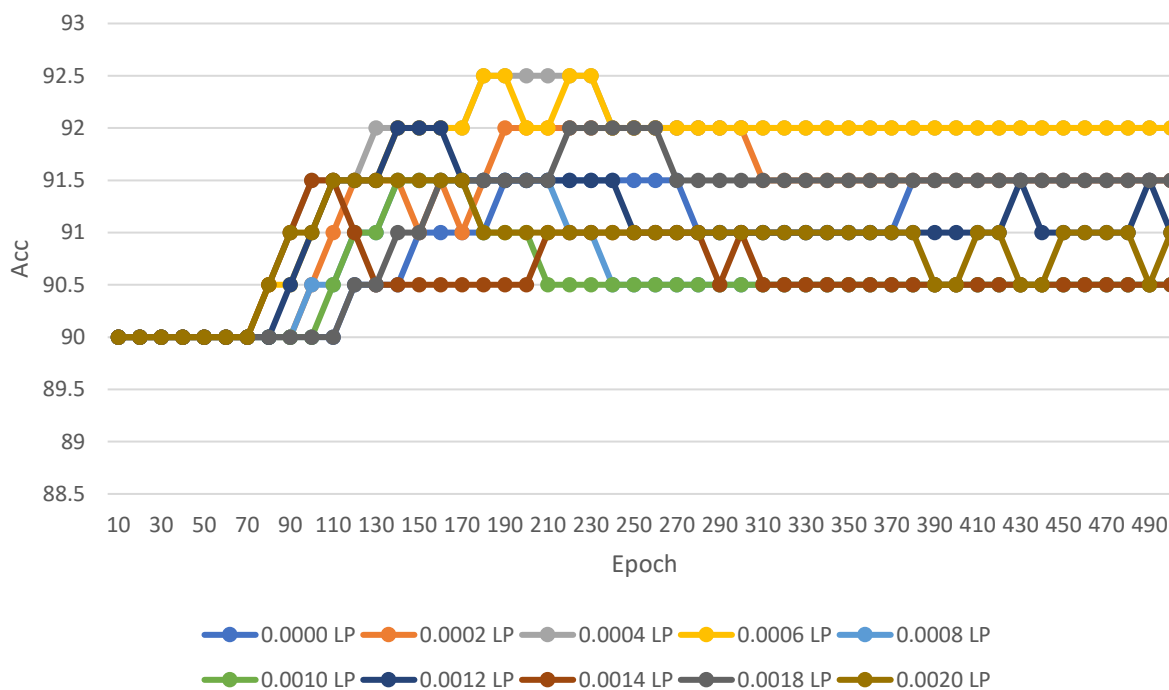
Epoch vs Accuracy of different L2 Penalty Value in test 3 using train set 1



Epoch vs Accuracy of different L2 Penalty Value in test 1 using train set 2



Epoch vs Accuracy of different L2 Penalty Value in test 2 using train set 2



Epoch vs Accuracy of different L2 Penalty Value in test 3 using train set 2

