Assignment 1 Writeup

Name: Willie Zeng
GT Email: wzeng40@gatech.edu

Two-Layer Neural Network

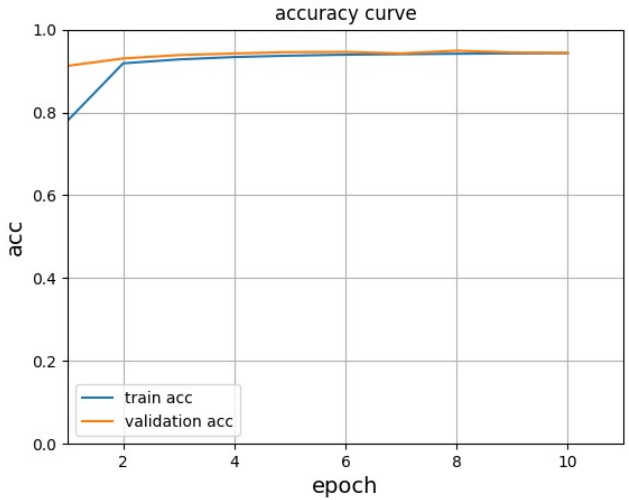
1. Learning Rates

Tune the learning rate of the model with all other default hyper-parameters fixed. Fill in the table below:

	lr=1	lr=1e-1	lr=5e-2	lr=1e-2
Training Accuracy	0.9453	0.9251	0.9145	0.762
Test Accuracy	0.9504	0.9255	0.9142	0.7579

1. Learning Curve

Plot the learning curves using the learning rates from the previous slide and put them below (you may add additional slides if needed).



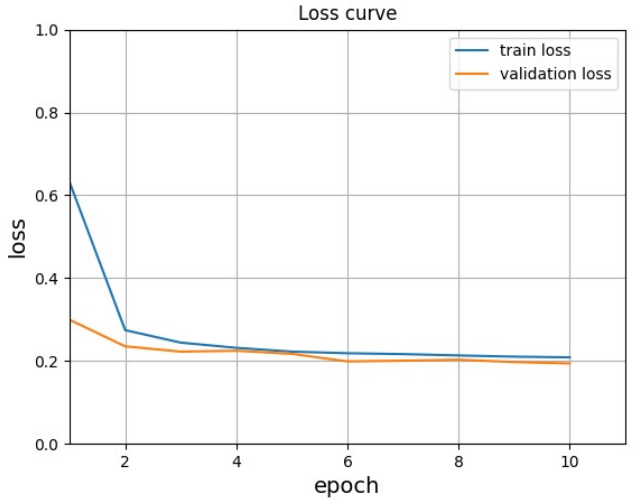
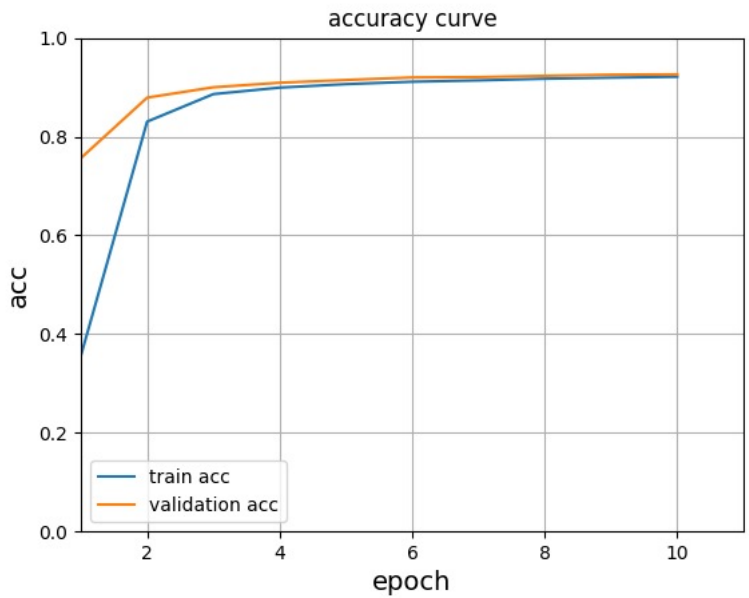


Figure 1: learning rate= 1

Learning Curve



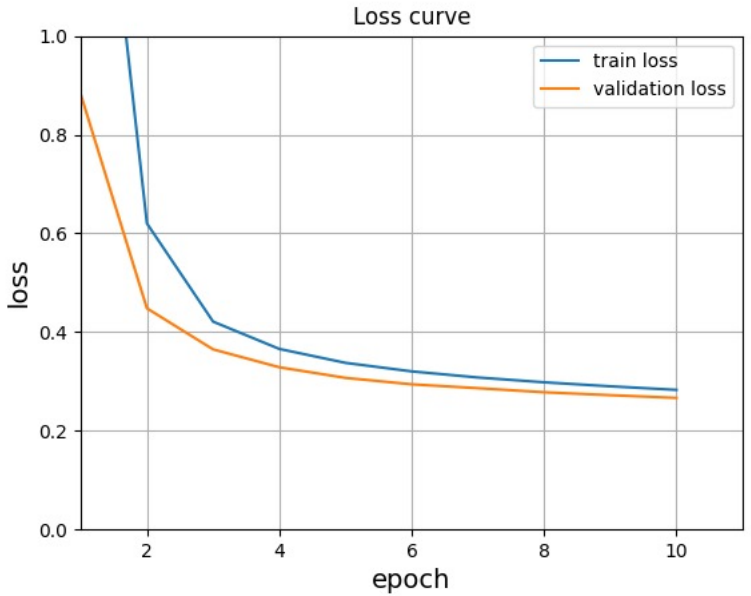
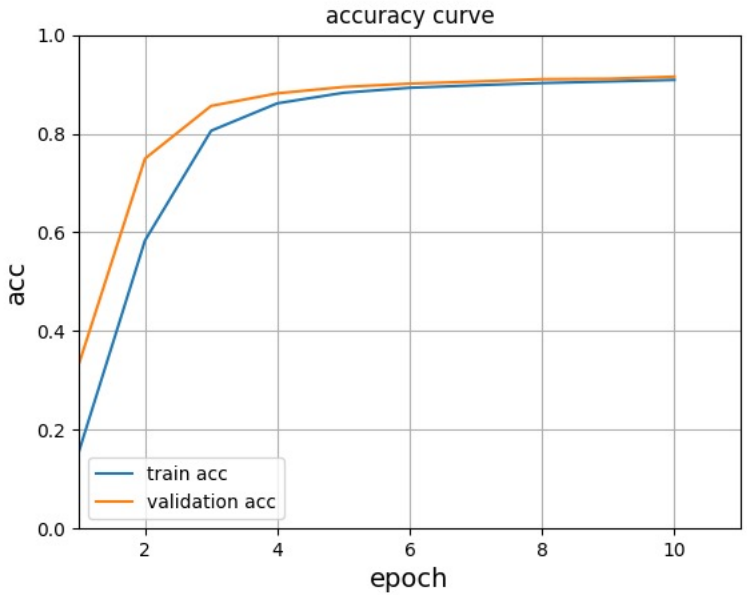


Figure 2: lr=1e-1

Learning Curve



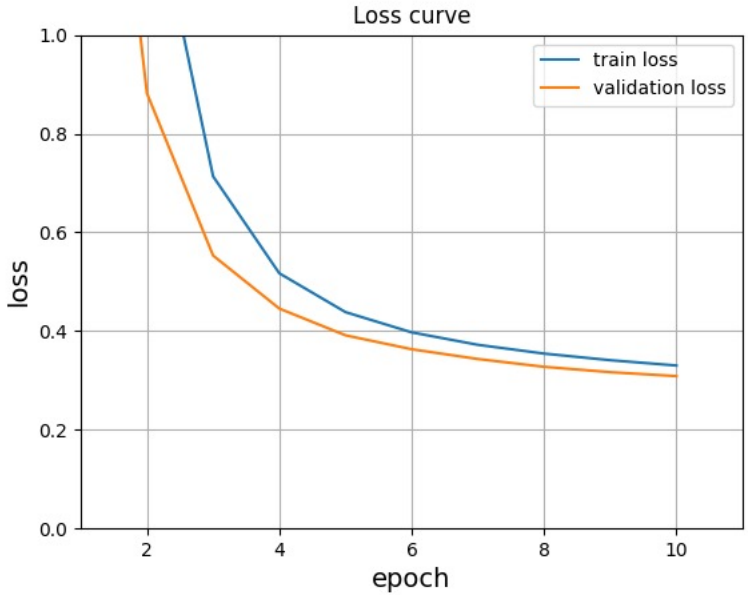
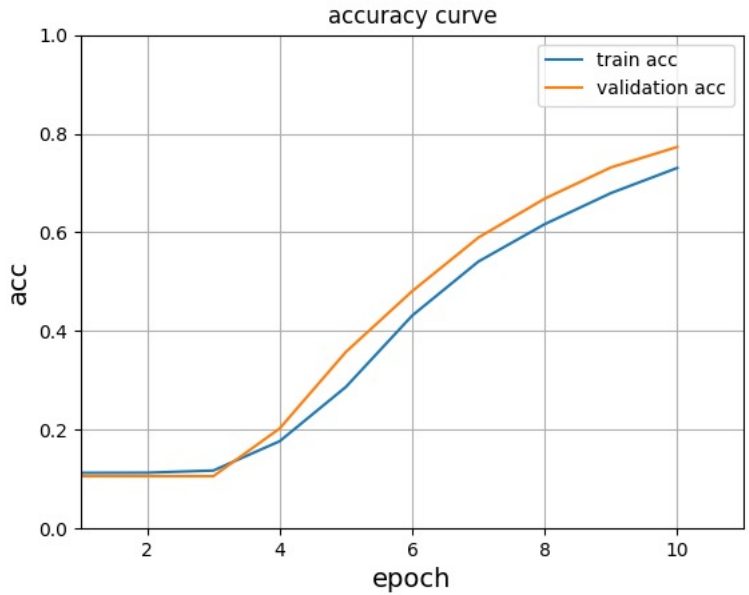


Figure 3: lr=5e-2

Learning Curve



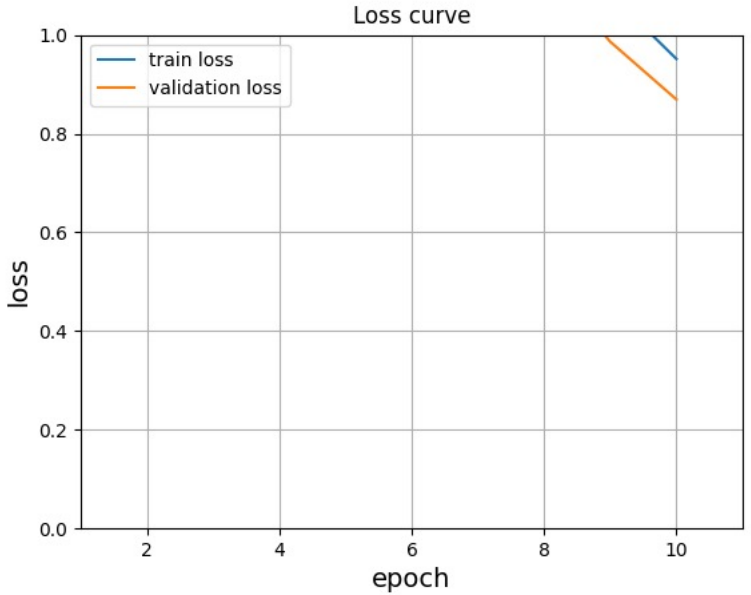


Figure 4: lr=1e-2

1. Learning Rates

As the Learning rate decreases from 1 to e^-2, the test accuracy and train accuracy also decreases. This means that for our default epoch of 10, the learning rate should be at least 1 (to achieve optimal test results), lower learning rates would require more epochs as it updates the weights slower. Also, the loss for each epoch decreases at a slower rate while remaining high (0.75 – 1), which is another indicator that the learning rate is too low for 10 epochs.

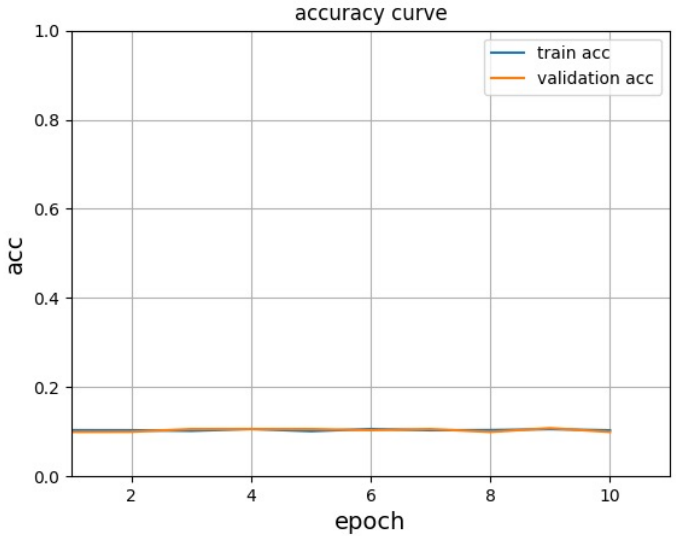
2. Regularization

Tune the regularization coefficient of the model with all other default hyper-parameters fixed. Fill in the table below:

	alpha=1	alpha=1e-1	alpha=1e-2	alpha=1e-3	alpha=1e-4
Training Accuracy	0.1039	0.3340	0.8845	0.9213	0.9314
Validation Accuracy	0.979	0.3663	0.8955	0.9282	0.9321
Test Accuracy	0.1029	0.3674	0.8934	0.9257	0.9323

2. Regularization

Plot the learning curves using the regularization coefficients from the previous slide and put them below (you may add additional slides if needed).



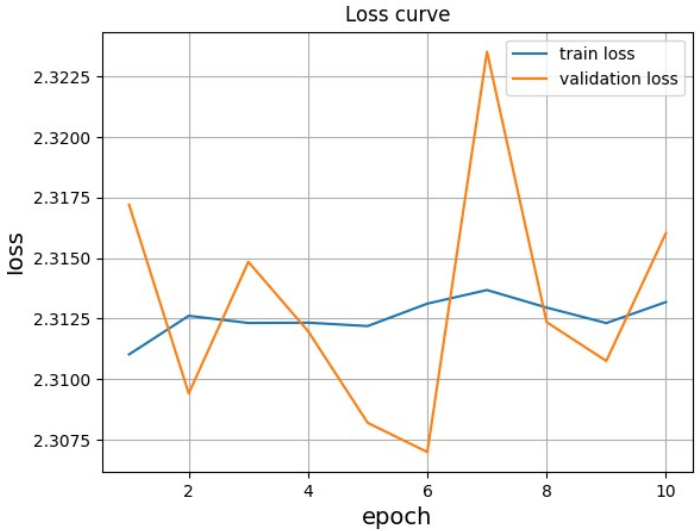


Figure 5: alpha=1

Regularization

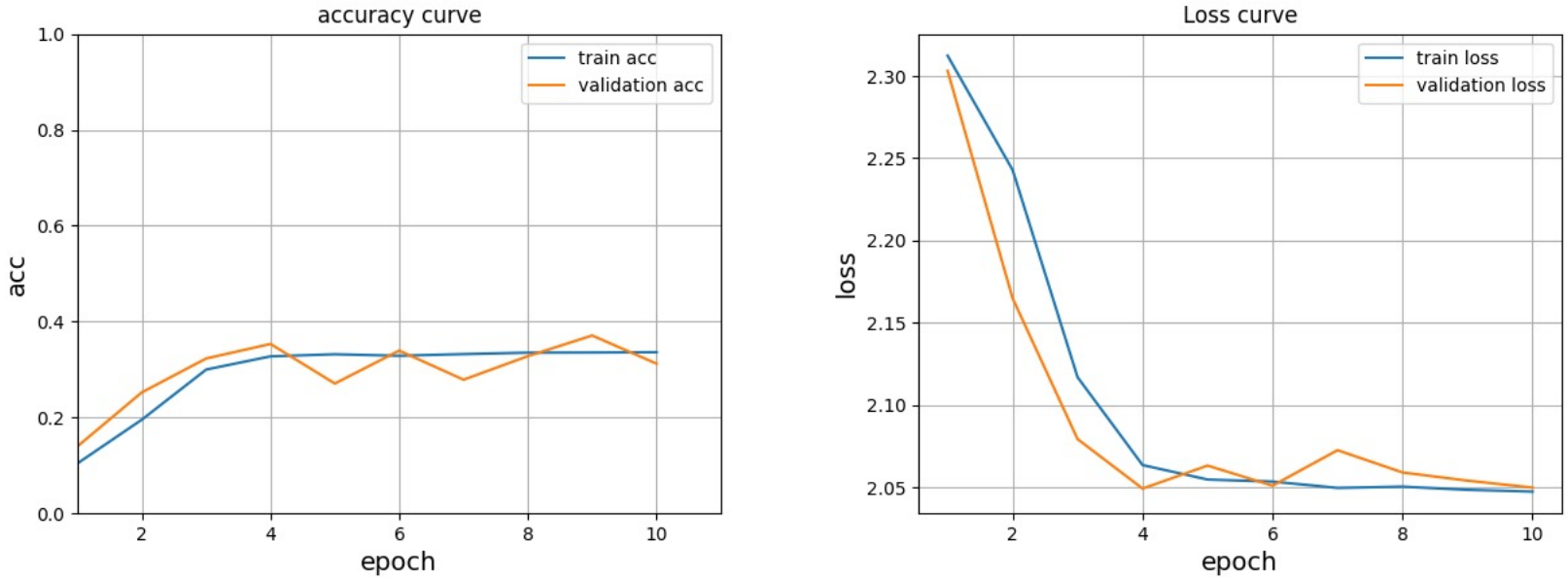
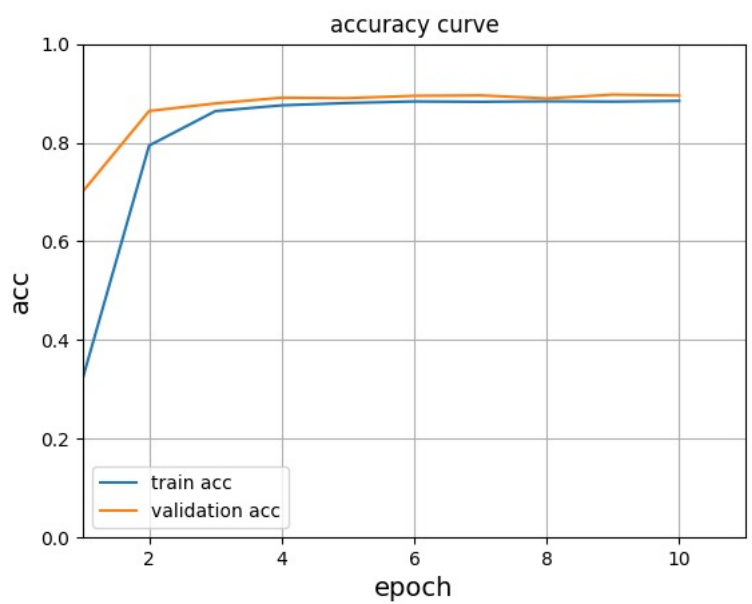


Figure 6: alpha=1e^-1

Regularization



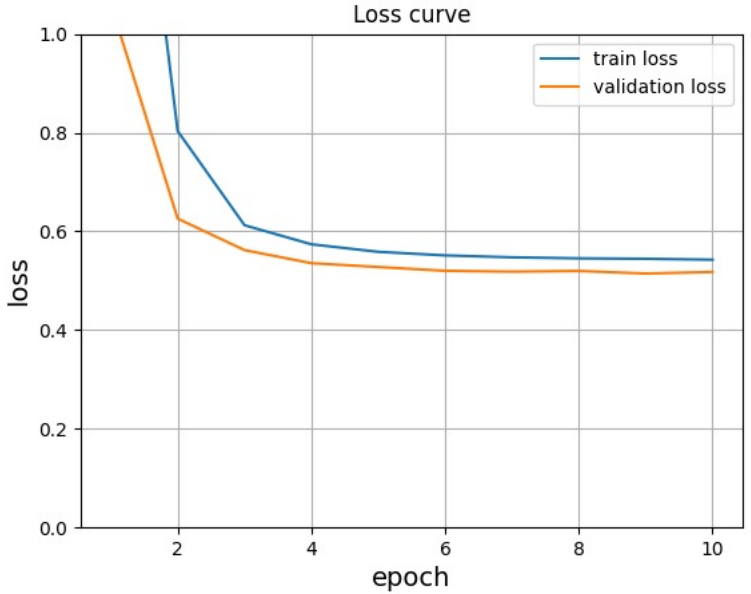


Figure 7: alpha=1e^-2

Regularization

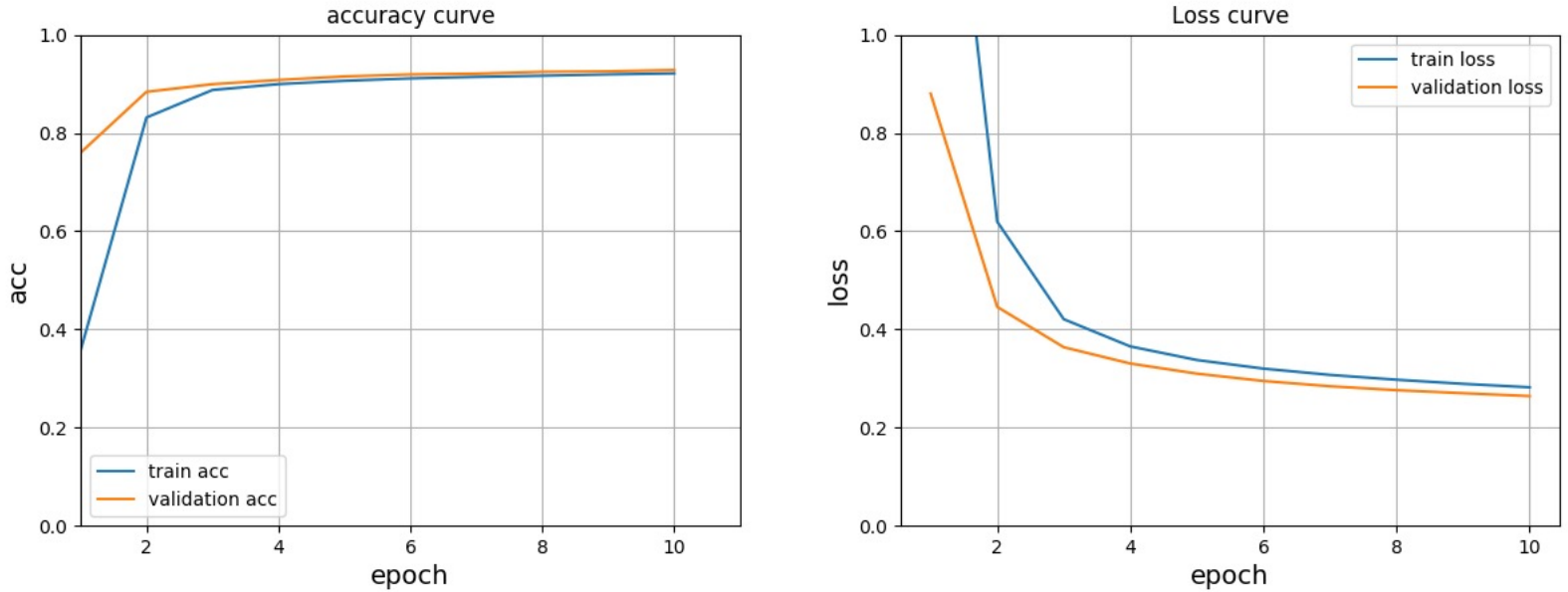
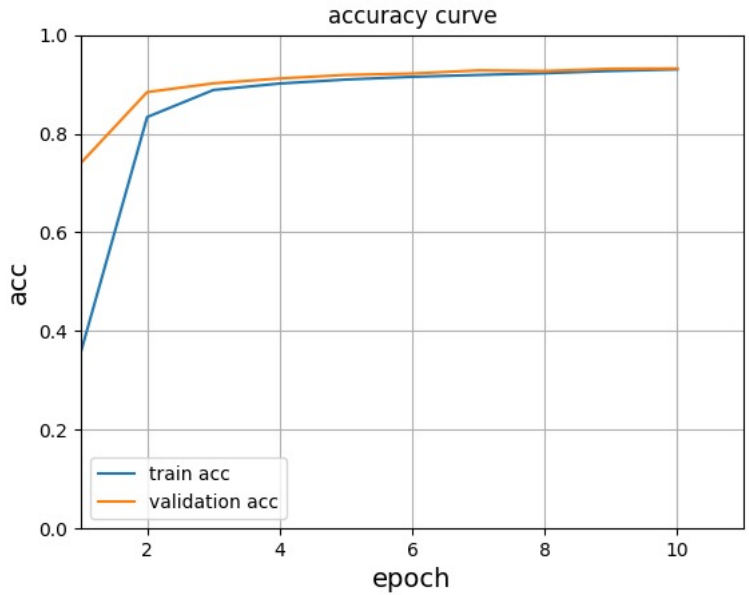


Figure 8: alpha=1e^-3

Regularization



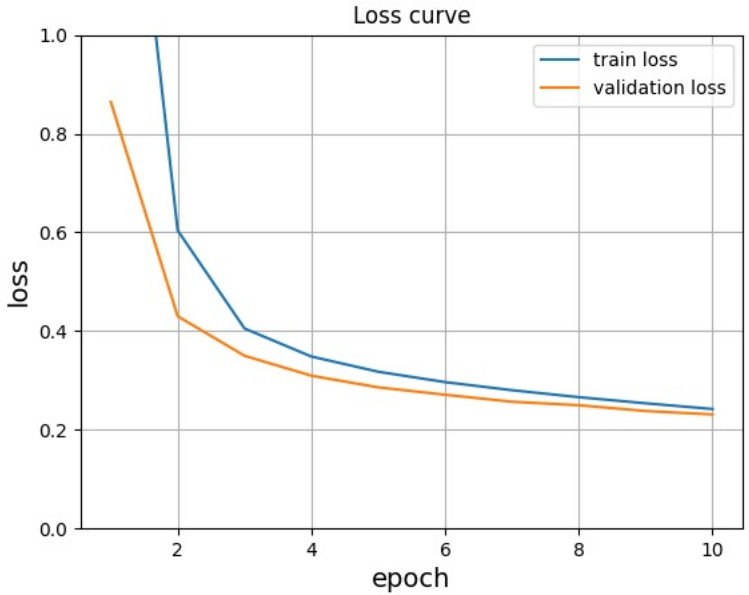


Figure 9: alpha=1e^-4

2. Regularization

Describe and Explain your findings:

As Alpha decreases from 1 to 1e-4, the accuracy increases. The whole point is to to find the balance between underfitting and overfitting by finding out how much should we **constrain the weights**.

Increasing the Alpha should be done in response to overfitting because it smooths the accuracy result by encouraging smaller weights. However, increasing Alpha too much (in this case, Alpha = 1) will result in a very slow learning process for just 10 epochs because the weights are too small.

Decreasing the Alpha should be done in response to underfitting because it encourages larger weights. However, in many instances, decreases the alpha too much may give high variances. In this case, we decrease to an optimal amount as the accuracy results are still steadily/smoothly increasing.

3. Hyper-parameter Tuning

You are now free to tune any hyper-parameters for better accuracy. Create a table below and put the configuration of your best model and accuracy into the table:

Batch Size	Learning Rate	Regularization	Epochs	Momentum	**TEST ACCURACY**	**VALIDATION ACCURACY**	**TRAIN ACCURACY**
64	1	0.0001	12	0.9	0.9721	0.9712	0.9802

Explain why your choice works:
Based on our isolated experimentations, I've noticed that 0.0001 yields the best accuracy and learning rate 1 yields the best accuracy. I've then added more epochs to increase the number of trained iterations from 10 to 12 to allow the weights more time to find their optimal value.

3. Hyper-parameter Tuning

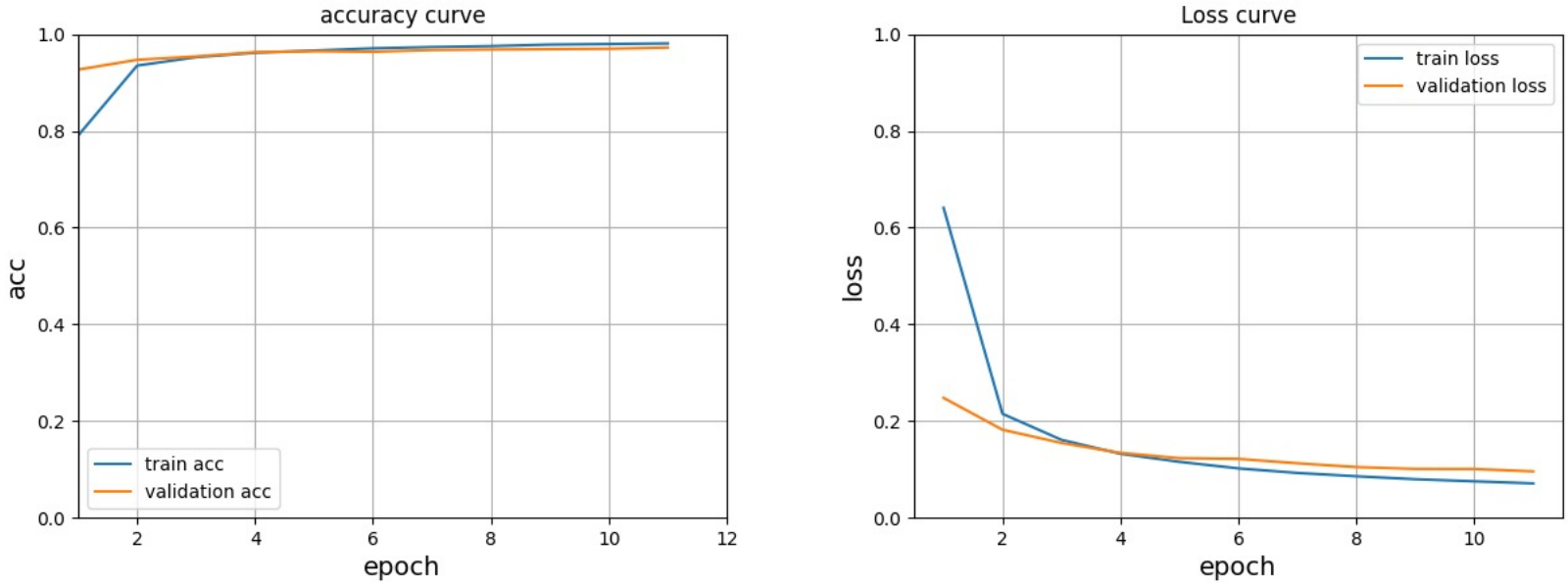


Figure 10: alpha=1e^-4