DD2424 Deep Learning for Data Science

Assignment 2

Report

Prashant Yadava

Code has been submitted at Canvas in an interactive python notebook format

Checking Analytic Gradient

The analytic gradient computations, when compared against the numerical gradient computations on a subset of real dataset were found to be negligible.

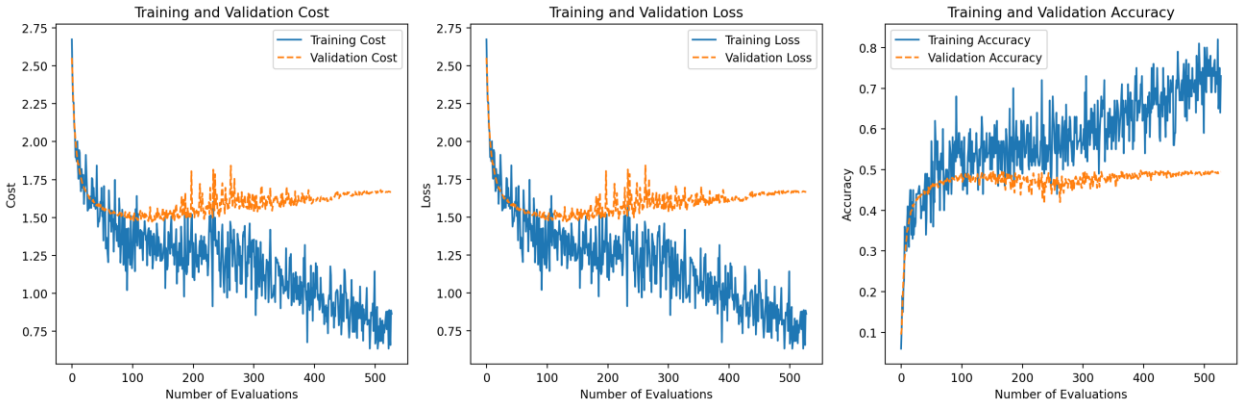
Results are as follows:

```
Shape: X_small: (20, 100), Y_small: (10, 100)
Shape: P: (10, 100)
Relative difference for W1: 4.0415370017064554e-07
Relative difference for b1: 3.120560786931215e-07
Relative difference for W2: 3.7594296587128416e-07
Relative difference for b2: 2.331878651081477e-07
```

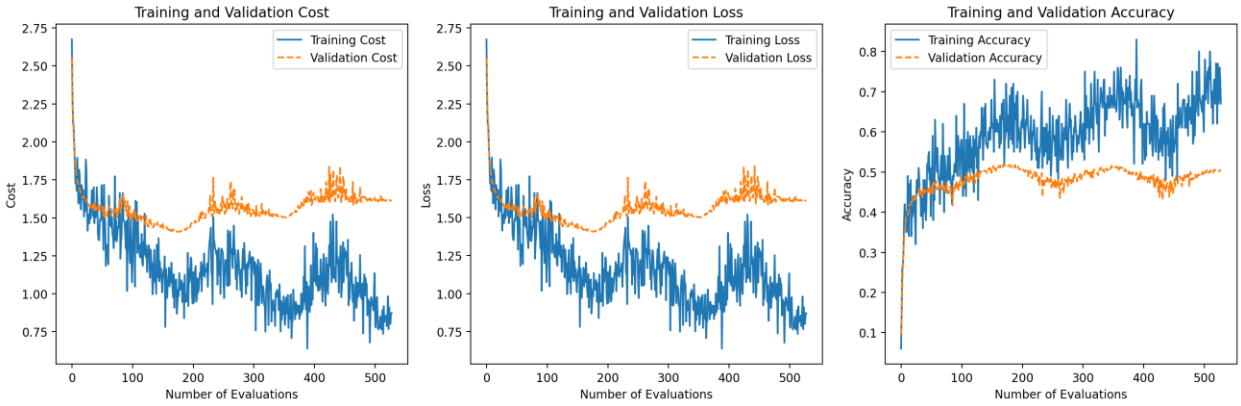
Table1: Differences between gradient computations between Analytical and Numerical gradient computations.

Curves for Training and Validation Loss

Plot 1: Training Curve for 1 cycle of training; Test accuracy was seen to be 45%



Plot 2: Training Curve for 3 cycles of training; Test accuracy was seen to be 46%



Cost:

Here the learning rate increases cyclically over the iterations. The costs initially decrease sharply, indicating that the model is learning and adjusting the weights effectively. As the learning rate decreases, the cost continues to decrease but at a slower rate.

Loss:

The loss follows a similar pattern to the cost, decreasing initially then stabilizing. The oscillations in the loss curve might be observed due to varying learning rate.

The loss is typically lower at the end of the cycle compared to the beginning, indicating learning progress.

Accuracy:

Accuracy typically increases initially as the model starts to learn. The fluctuations in the accuracy might correspond to the higher learning rate phases, but it generally trends upwards.

By the end of the cycle, the accuracy is seen to be higher than at the start, showing that the model improves.

Search for Lambda

Coarse Search

Hyper-Parameters:

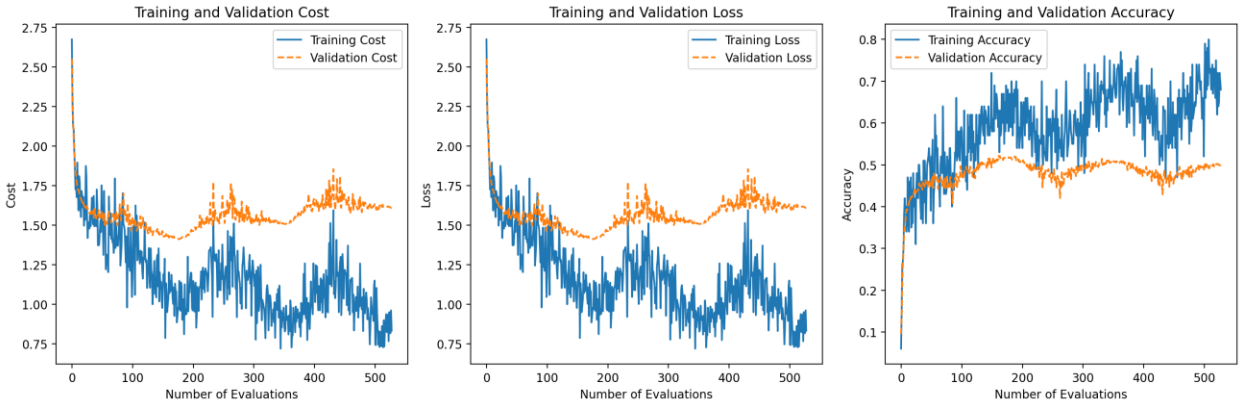
- ns = 900
- epoch = 48
- cycles = 3
- batch size = 100
- eta min = 1e-5
- eta max = 1e-1
- Lambda range: [1e-5, 0.001, 0.1]

Results:

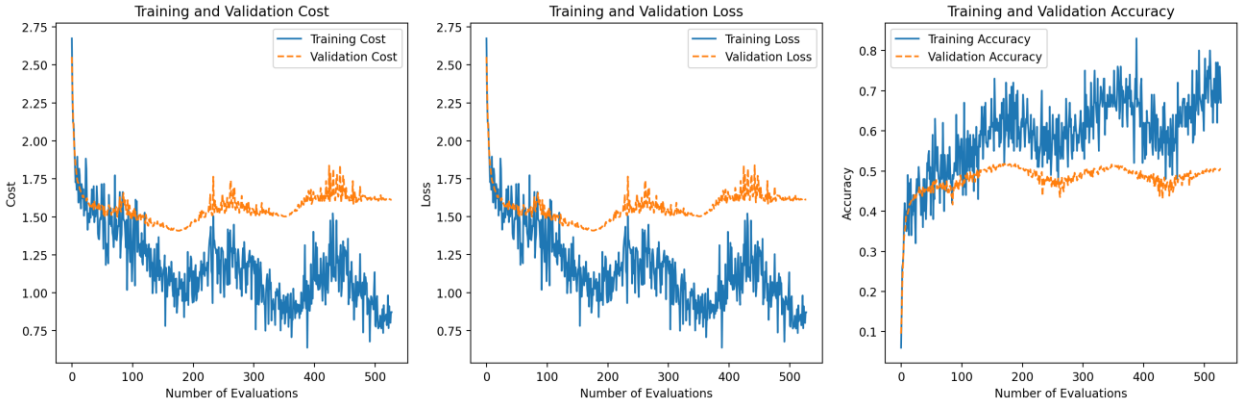
Lambda	Validation Accuracy
1e-05	0.5004
0.001	0.5046
0.1	0.5174

Corresponding curves:

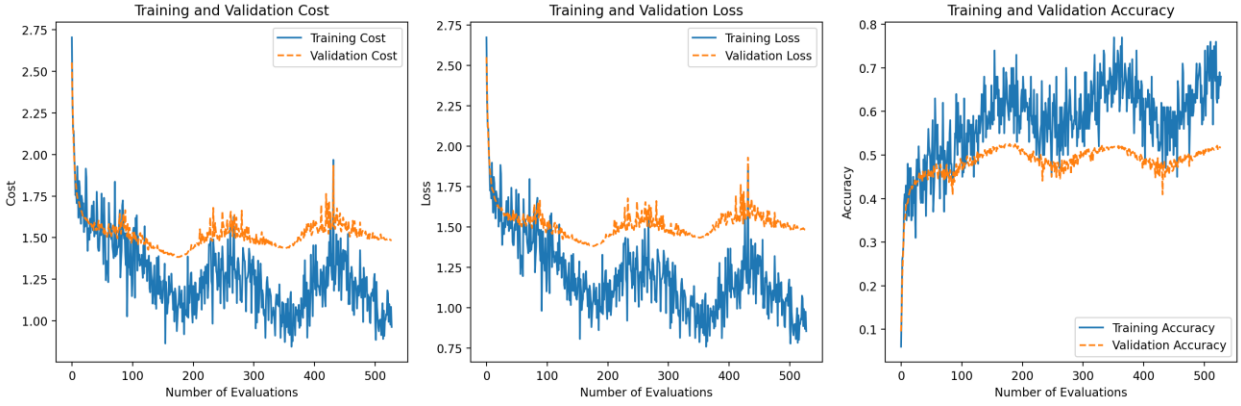
Plot 3: Lambda: 0.000010, Validation Accuracy: 0.5004



Plot 4: Lambda: 0.001000, Validation Accuracy: 0.5046



Plot 5: Lambda: 0.100000, Validation Accuracy: 0.5174



Fine Search

Hyper-Parameters:

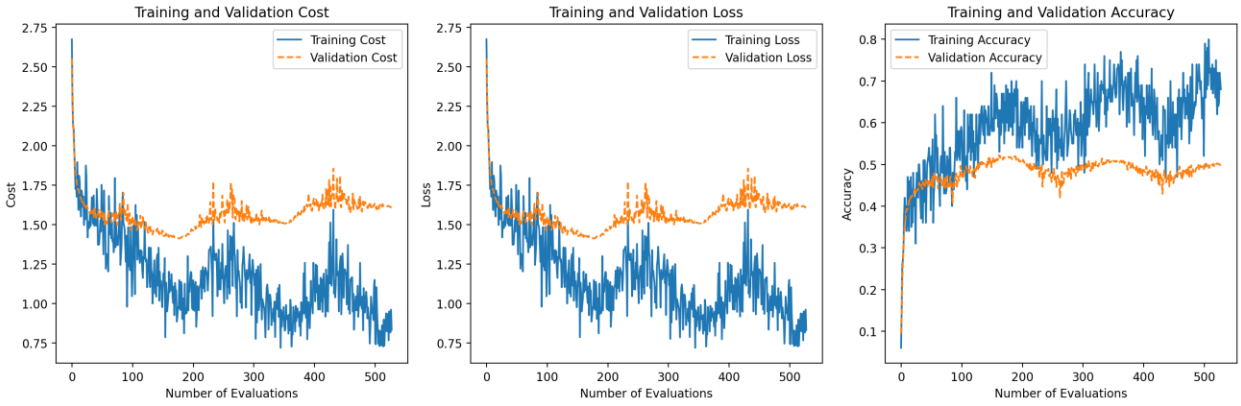
- ns = 900
- epoch = 48
- cycles = 3
- batch size = 100
- eta min = 1e-5
- eta max = 1e-1
- Lambda range: [1e-5, 0.001, 0.1]

Results:

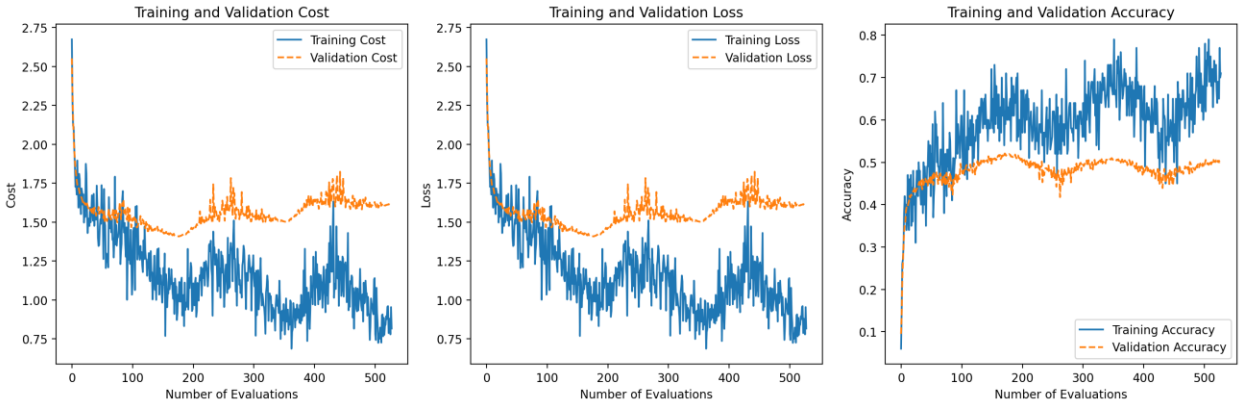
Lambda	Validation Accuracy
1e-05	0.5004
0.000037	0.4996
0.000139	0.4972
0.000518	0.5000
0.001931	0.4982
0.007197	0.5062
0.026827	0.5058
0.100000	0.5174

Corresponding curves:

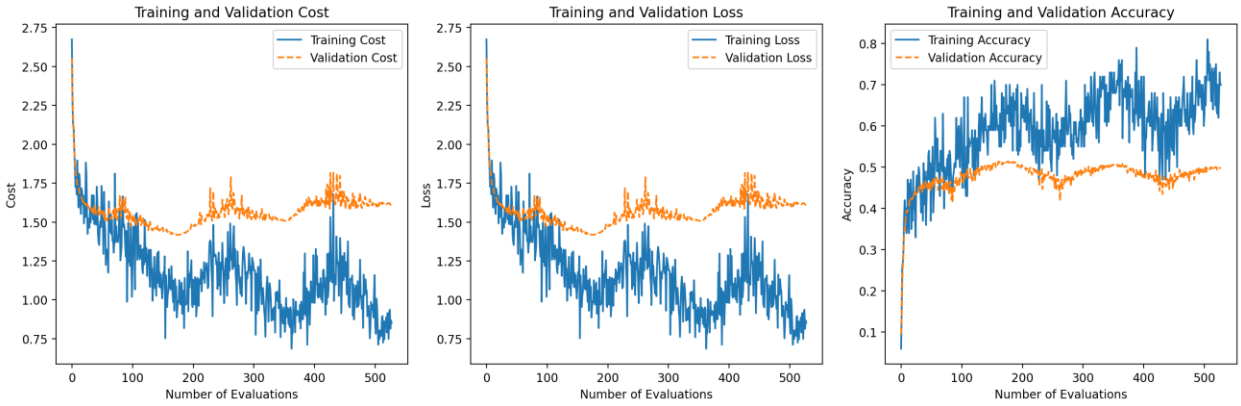
Plot: Lambda: 0.000010, Validation Accuracy: 0.5004



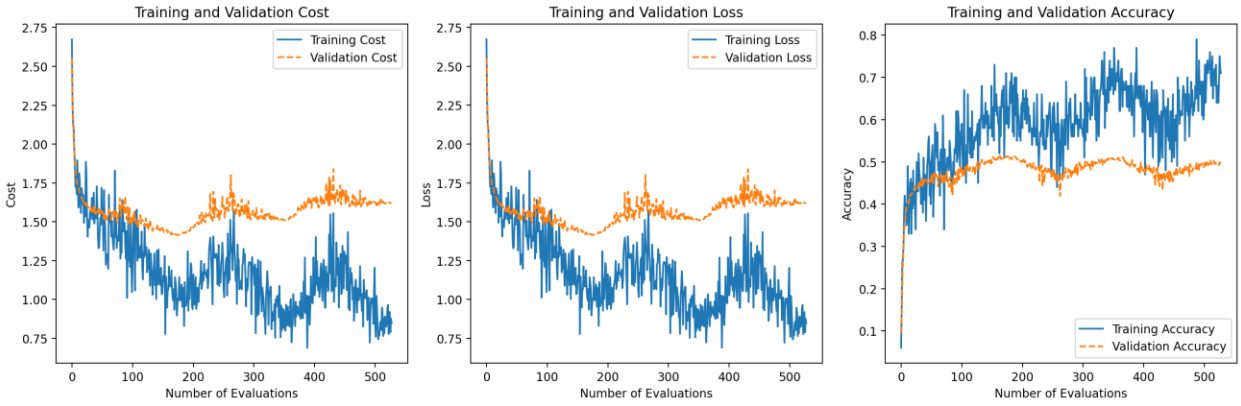
Plot: Lambda: 0.000037, Validation Accuracy: 0.4996



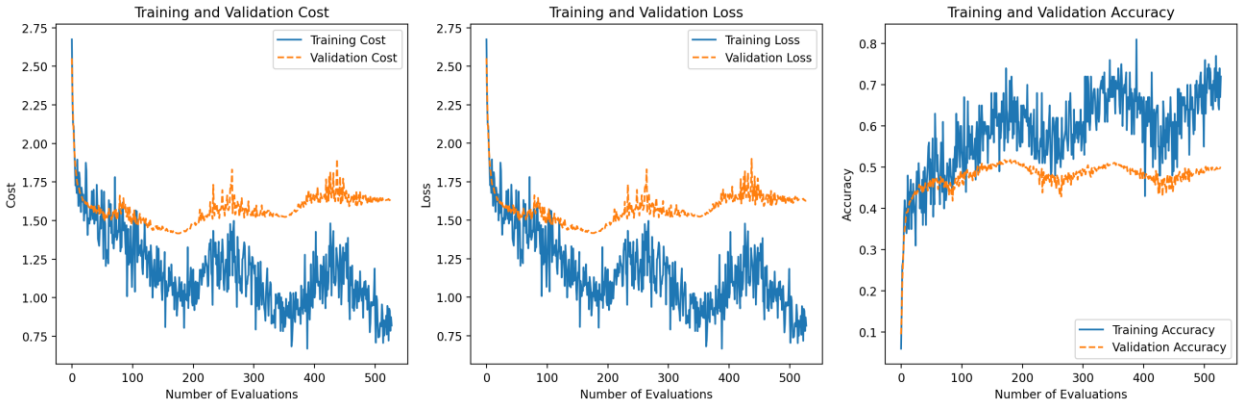
Plot: Lambda: 0.000139, Validation Accuracy: 0.4972



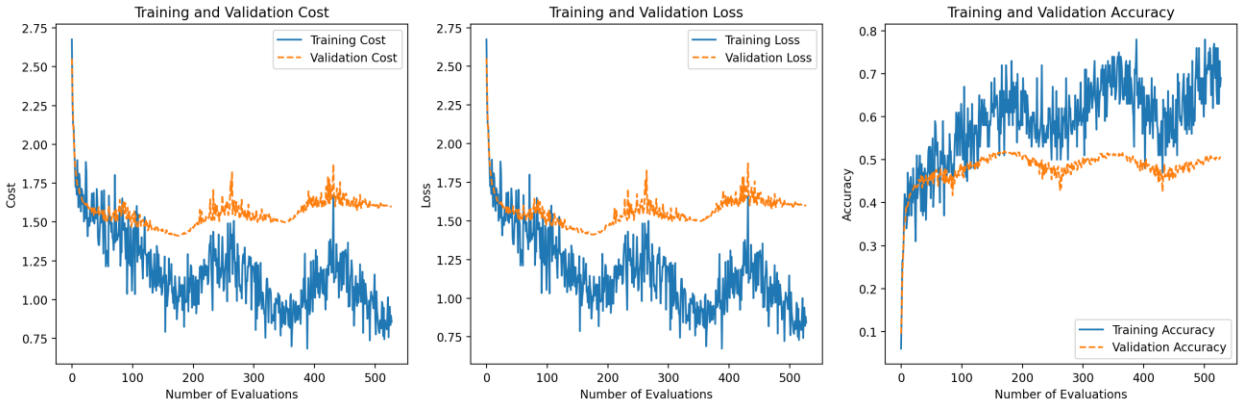
Plot: Lambda: 0.000518, Validation Accuracy: 0.5000



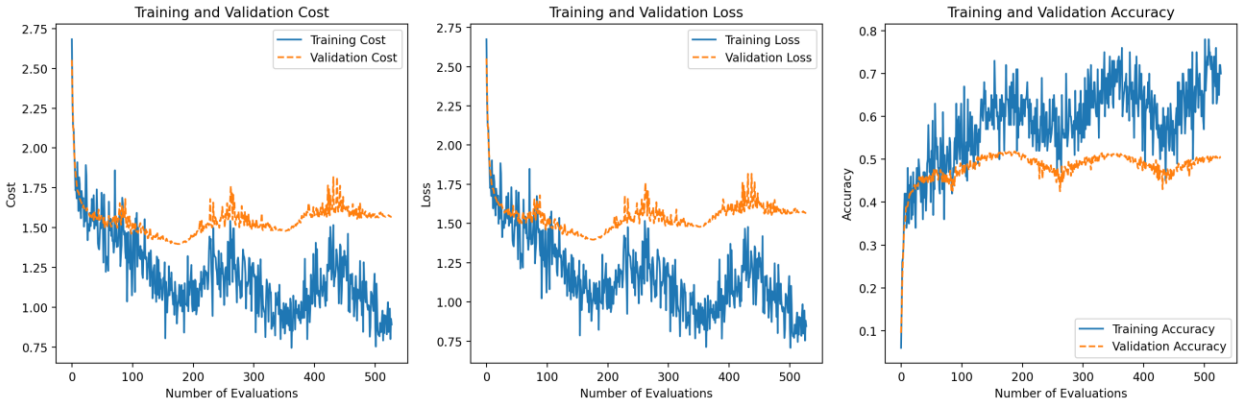
Plot: Lambda: 0.001931, Validation Accuracy: 0.4982



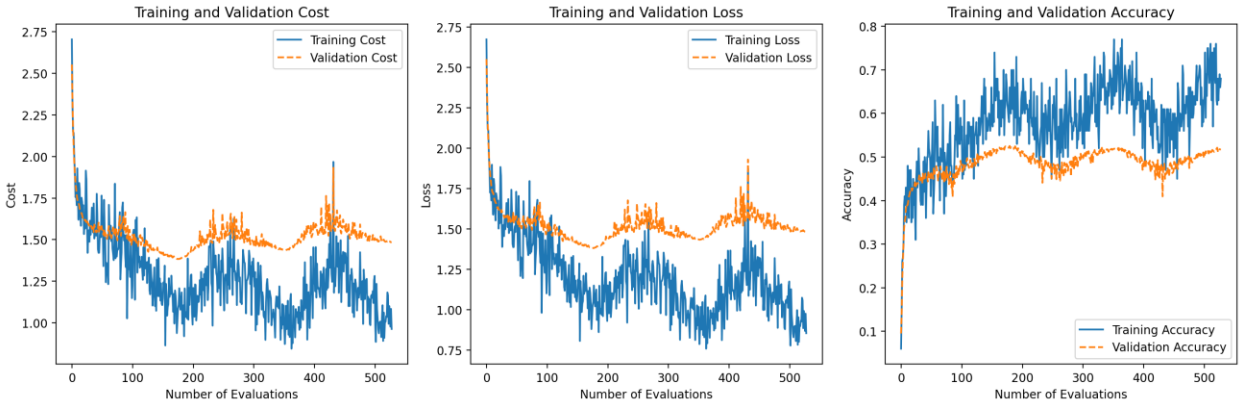
Plot: Lambda: 0.007197, Validation Accuracy: 0.5062



Plot: Lambda: 0.026827, Validation Accuracy: 0.5058



Plot: Lambda: 0.100000, Validation Accuracy: 0.5174



Training with best 3 lambdas on test data

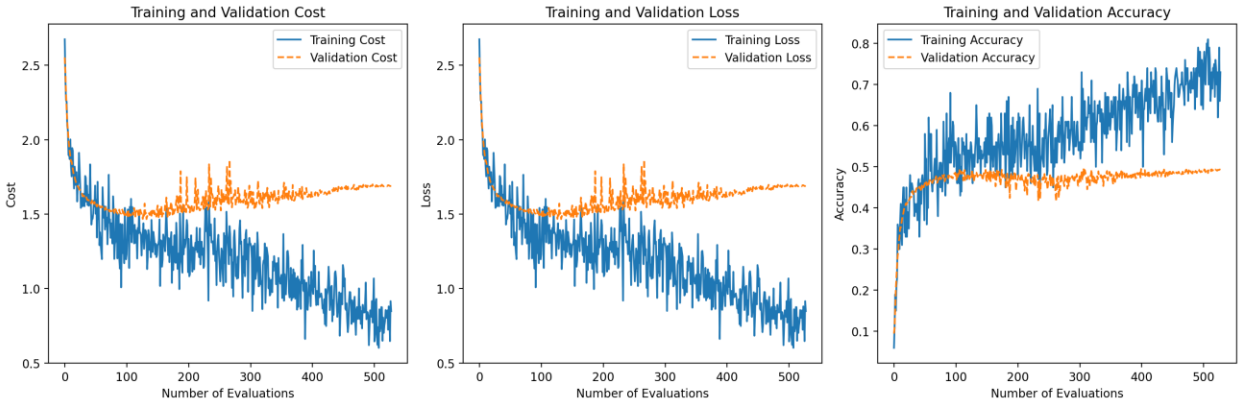
We chose the following 3 lambda values: [1e-5, 0.000037, 0.1]

After the training, the test accuracies with the lambdas were seen to be:

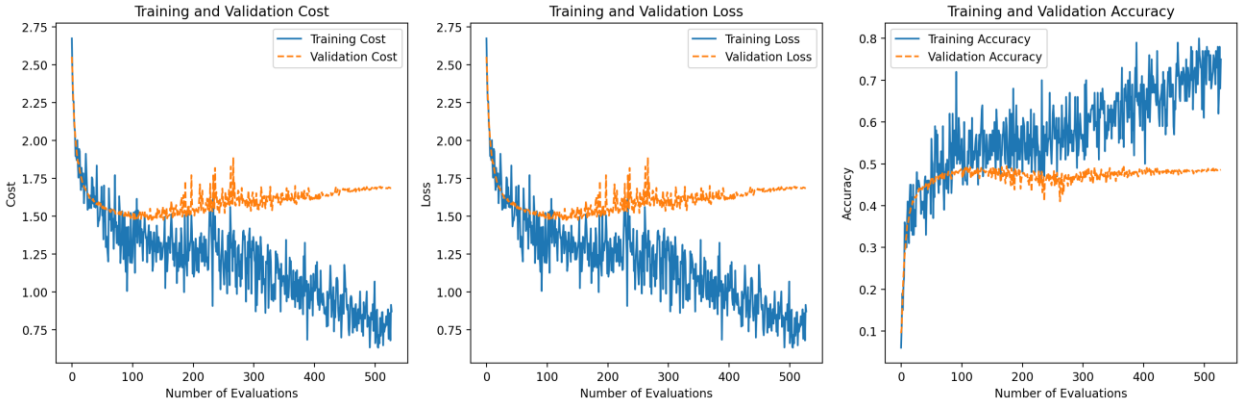
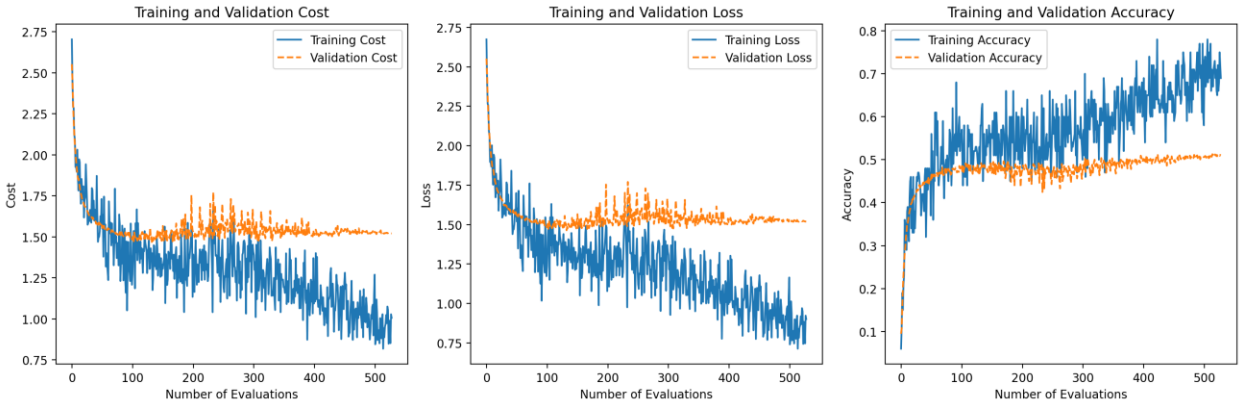
Lambda	**Test** Accuracy
1e-5	0.4586
0.000037	0.4573
0.1	0.5174

The corresponding plot showing Training and validation accuracies follow:

Plot: Lambda 1e-5, Validation accuracy 0.4928



Plot: Lambda 0.000037, Validation accuracy 0.4854



Plot: Lambda 0.01, Validation accuracy 0.5104