

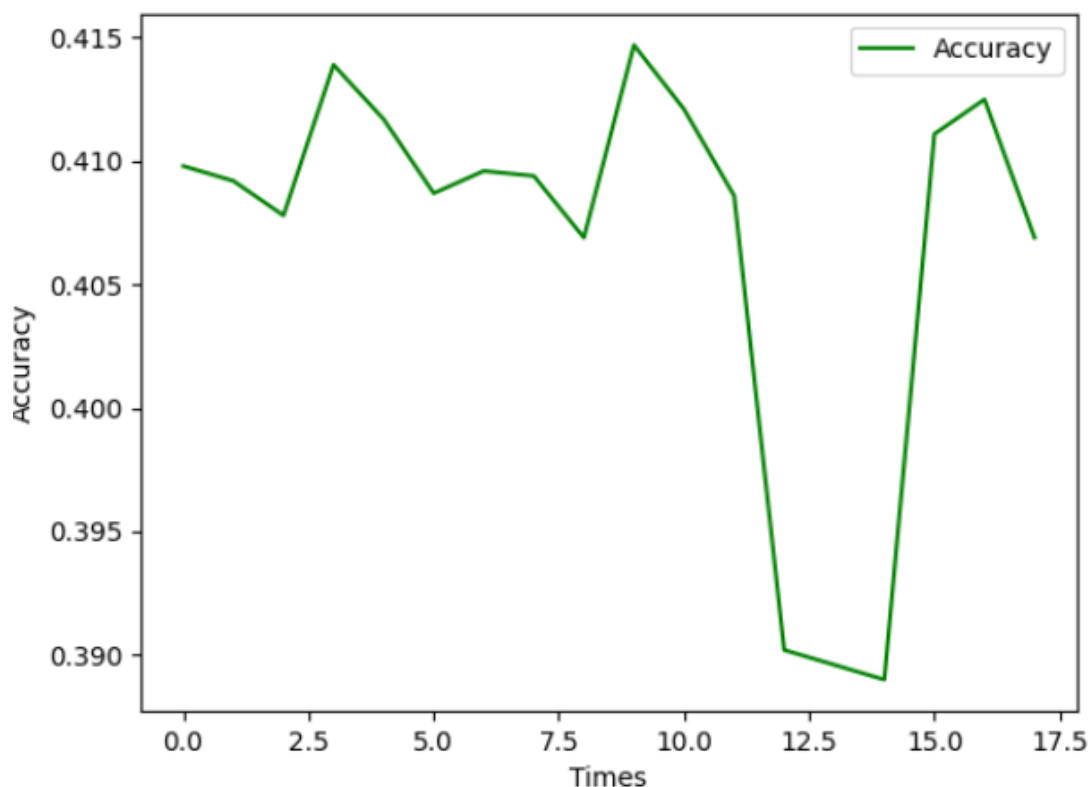
# Assignment1\_BonusPoint---Report

## Exercise 2.1:

According the tutorial, I adjust my code, including the adding use all data as training data and separate 1000 data as validation (174 ~ 192), enhancing the data (195 ~ 211) and doing grid searching (236 ~ 252) while finishing the learning rate decaying (123), the below sheet shows the hyperparameters which I used.

Batch_size	eta		lam
100	0.001	0.01	0.01
100			0.001
500			0.1

The result is a list: [0.4098, 0.4092, 0.4078, 0.4139, 0.4117, 0.4087, 0.4096, 0.4094, 0.4069, 0.4147, 0.4121, 0.4086, 0.3902, 0.3896, 0.389, 0.4111, 0.4125, 0.4069] and the chart shows the result list.

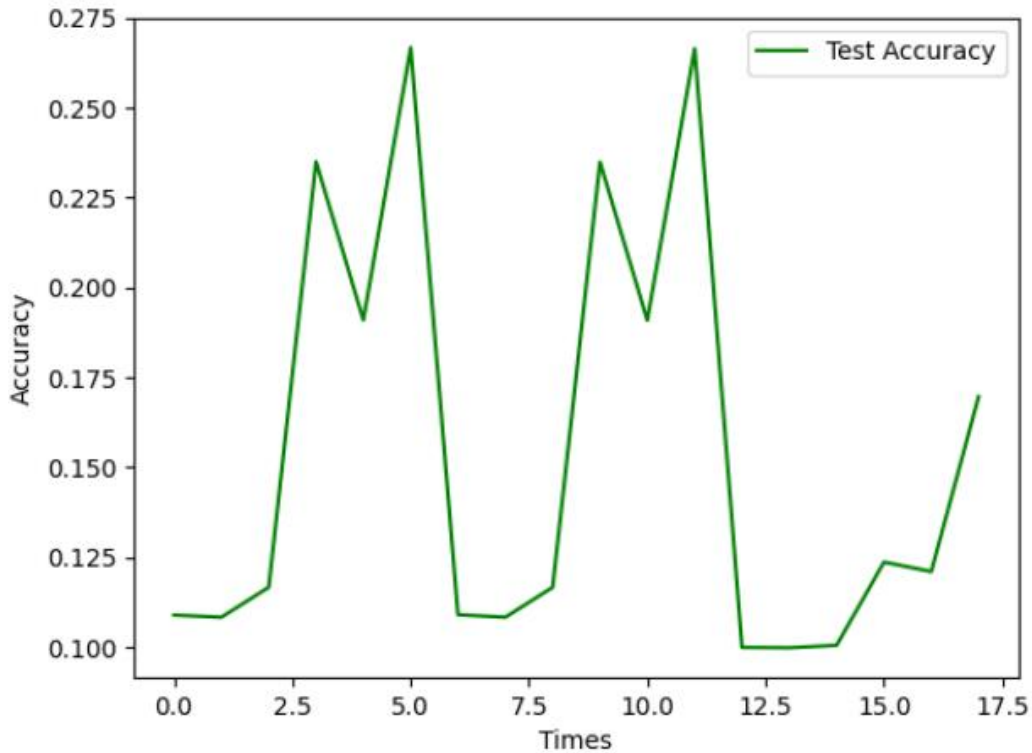


Obviously ,the **best hyperparameter** is batch\_size=100, eta=0.01, lam=0.01 and result is 0.4147

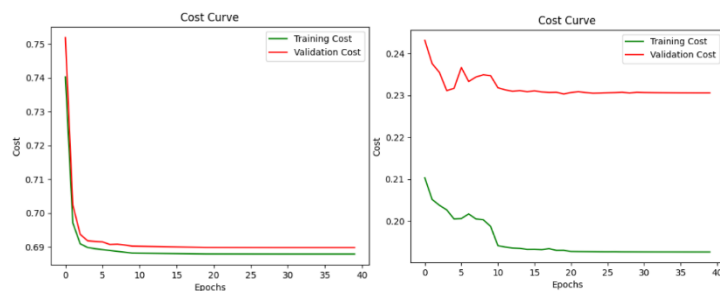
## Exercise 2.2:

When I changing the loss function and gradient function, I also keep to find the **best hyperparameter**, it is batch\_size=100, eta=0.01, lam=0.1 and the result is 0.2667. The below chart shows the accuracy.

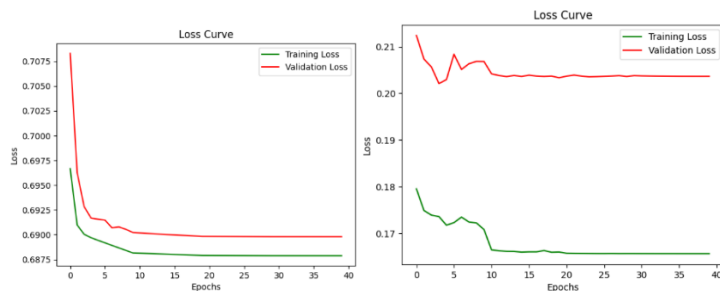
The accuracy list is: [0.109, 0.1084, 0.1167, 0.235, 0.191, 0.2667, 0.1091, 0.1084, 0.1167, 0.2349, 0.1909, 0.2664, 0.1, 0.0999, 0.1006, 0.1237, 0.1211, 0.1697]



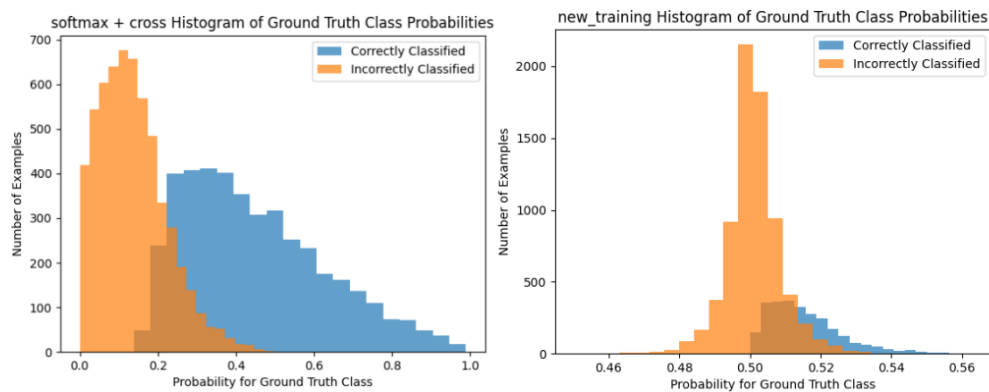
The below two charts shows the **cost** in new training procedure (left) and Softmax + cross-entropy(right):



The below two charts shows the **loss** in new training procedure (left) and Softmax + cross-entropy(right):



The **Histogram** of new training and softmax + cross is showed:



### Chart Analyzing:

According to the **histograms**, the **SoftMax + cross model** tends to push the true-class probability of correctly classified samples toward higher values, while incorrectly classified samples cluster at lower probabilities, showing a more distinct separation. In contrast, the **new\_training model** has most probabilities hovering around 0.5, causing a large overlap between correct and incorrect samples and indicating that the model does not assign very high or very low probabilities to the true class, thus exhibiting weaker separation overall.

In the **cost and loss curves** for SoftMax + cross, the model converges quickly and does not show obvious signs of overfitting. However, there is a noticeable gap between the training and validation curves, indicating that the model fits the training data better than it generalizes to the validation data, suggesting a certain degree of overfitting.

In New\_training, the performance on the training and validation sets is very close, which shows that the model does not exhibit obvious overfitting upon convergence and generalizes well.