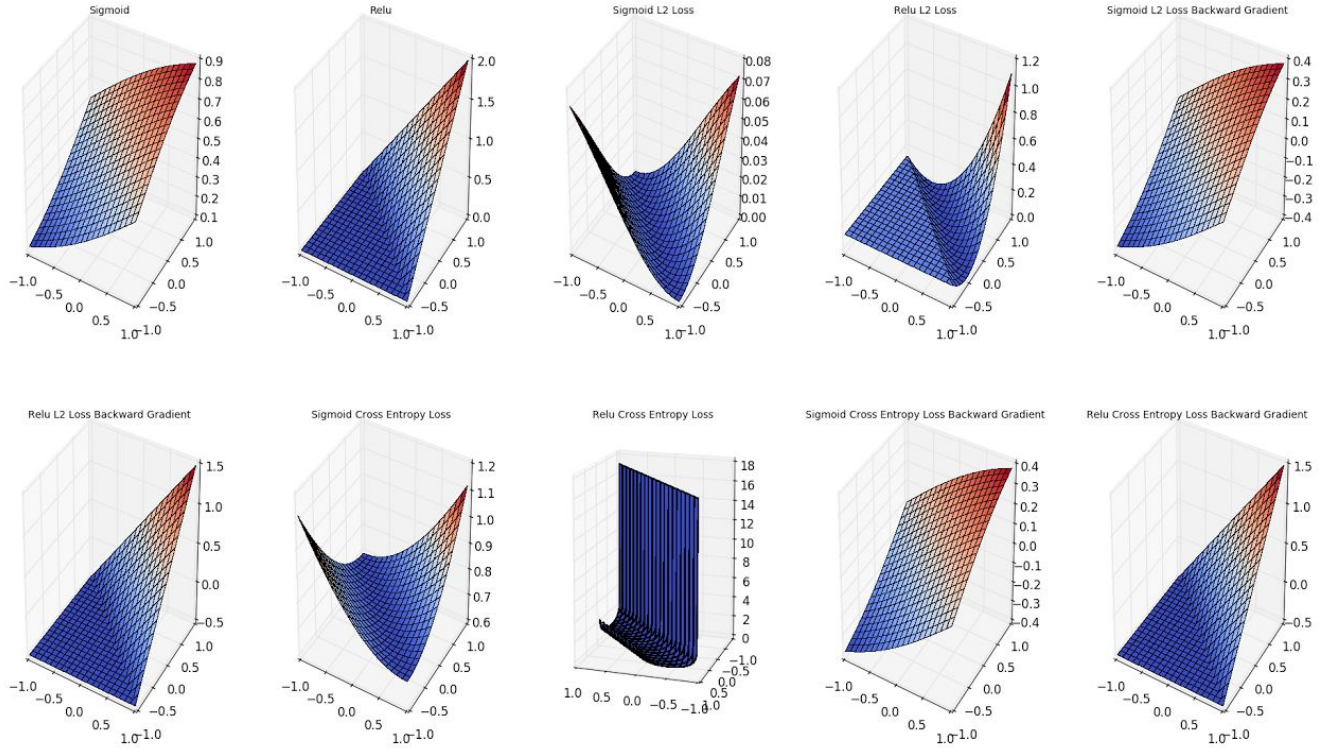
# PROJECT 4B REPORT

By Shraddha Jain, Tarmily Wen, Hongrui Zheng

**P1**




a) Cross entropy and L2 have different derivatives and different purposes. Cross entropy is prefered for classification, and L2 is prefered for regression.

b) The derivative of the L2 loss is:

$$\frac{\partial L}{\partial w} = (\sigma(z) - y)\sigma'(z)x$$

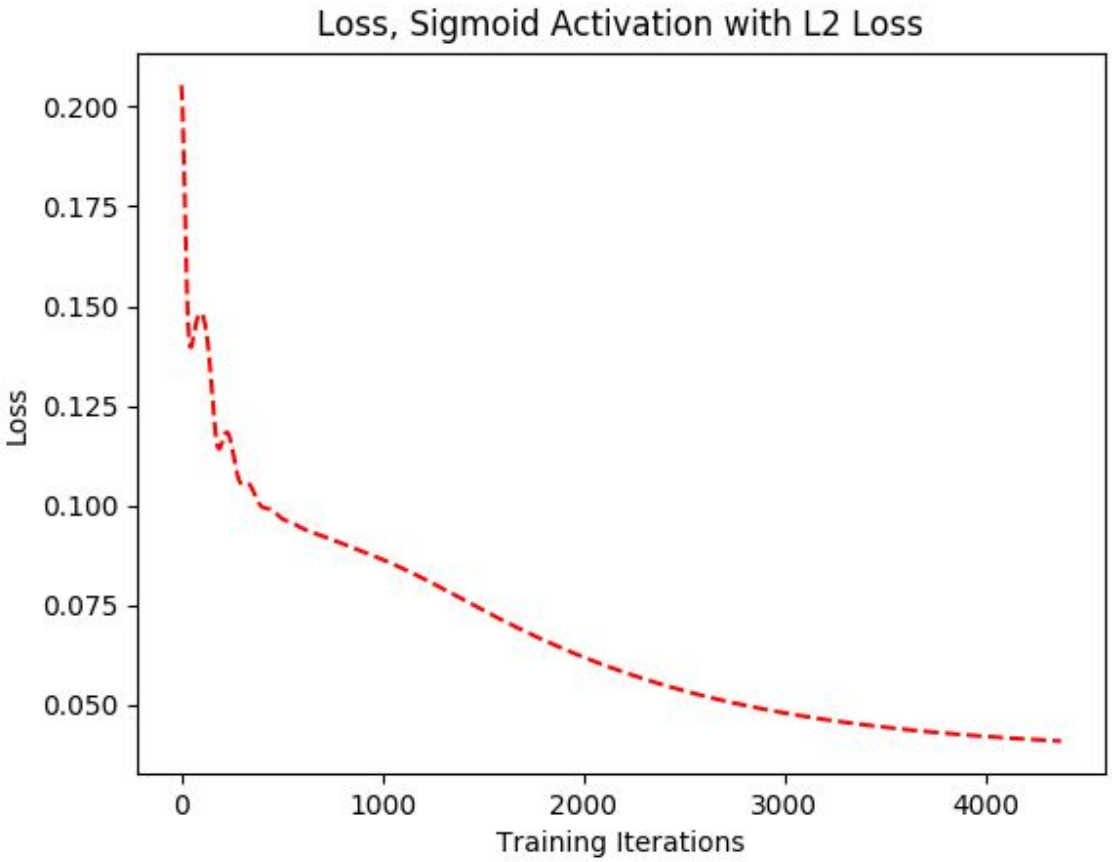
And the derivative of the Cross Entropy loss is:

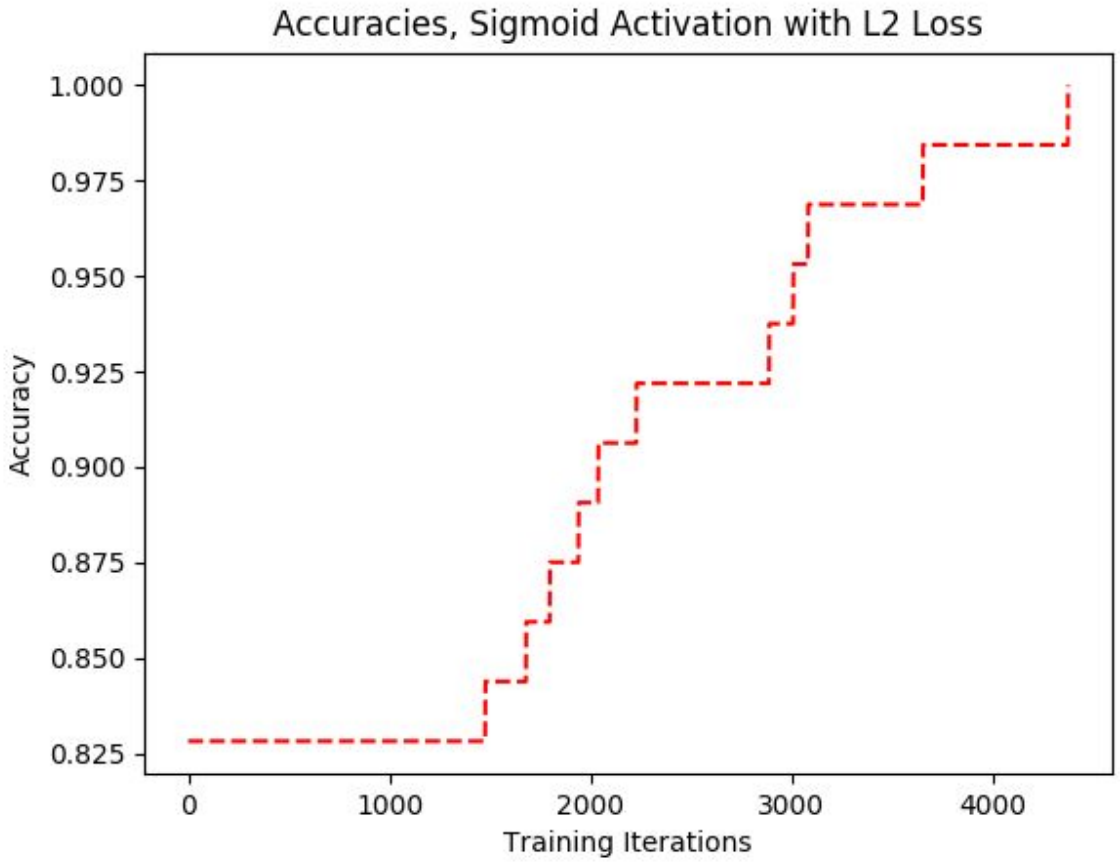
$$\frac{\partial L}{\partial w} = (\sigma(z) - y)x$$

Cross Entropy loss is designed such that the gradient doesn't contain the derivative of the activation function.

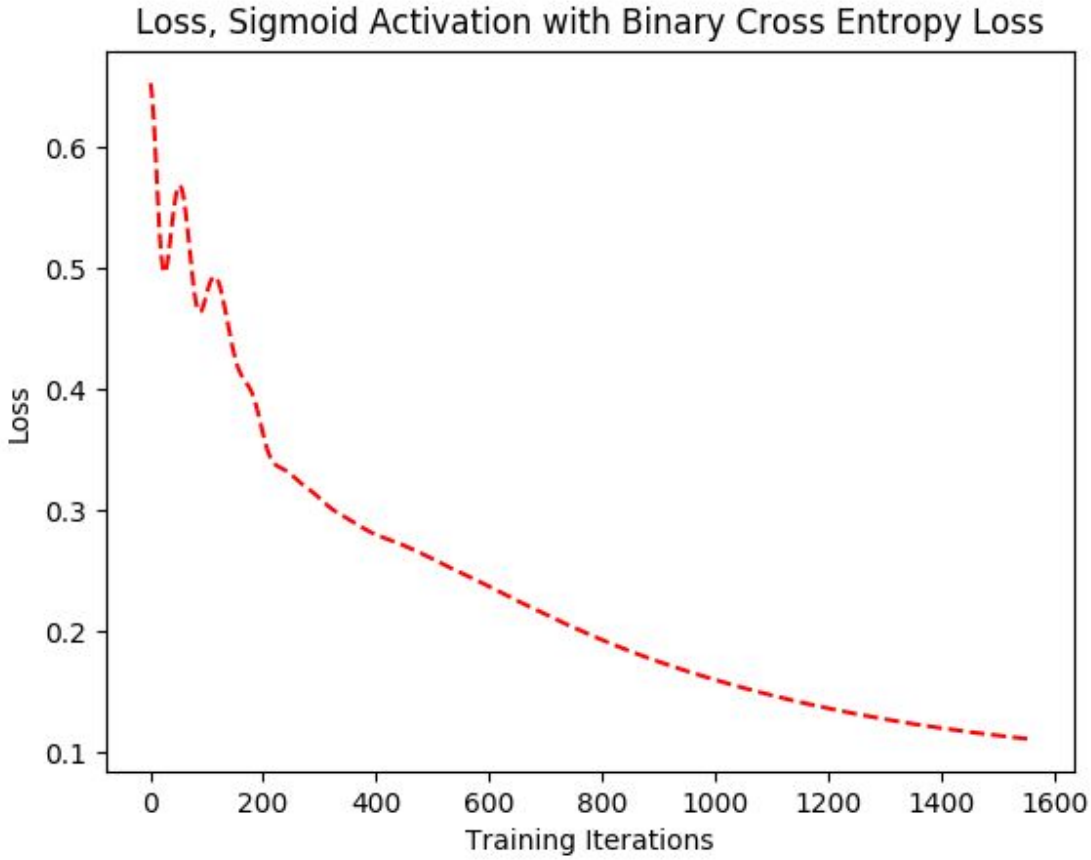
c) Since it's a log loss function, the loss is much higher when the prediction is further away from the ground truth, thus learning is much faster initially while using cross entropy loss.

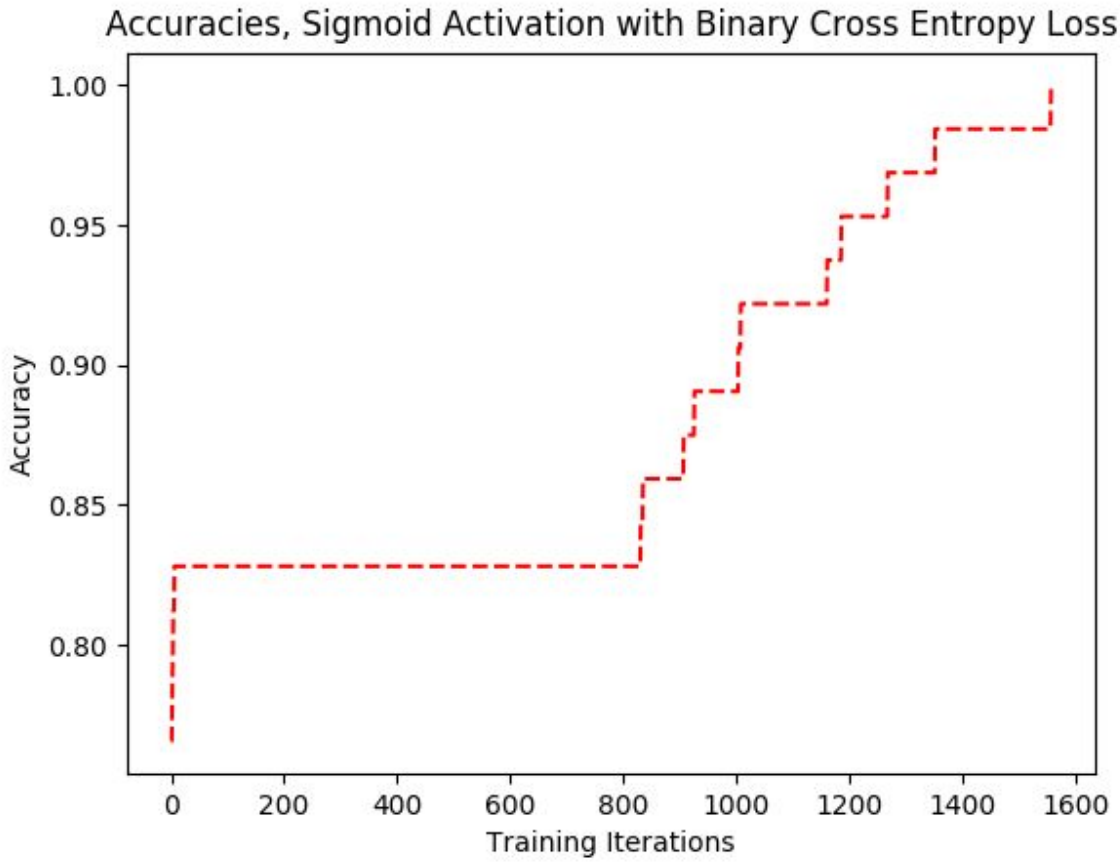
**P2.1.1:**


Loss, Sigmoid Activation with L2 Loss
Loss
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0
1000
2000
3000
4000
Training Iterations

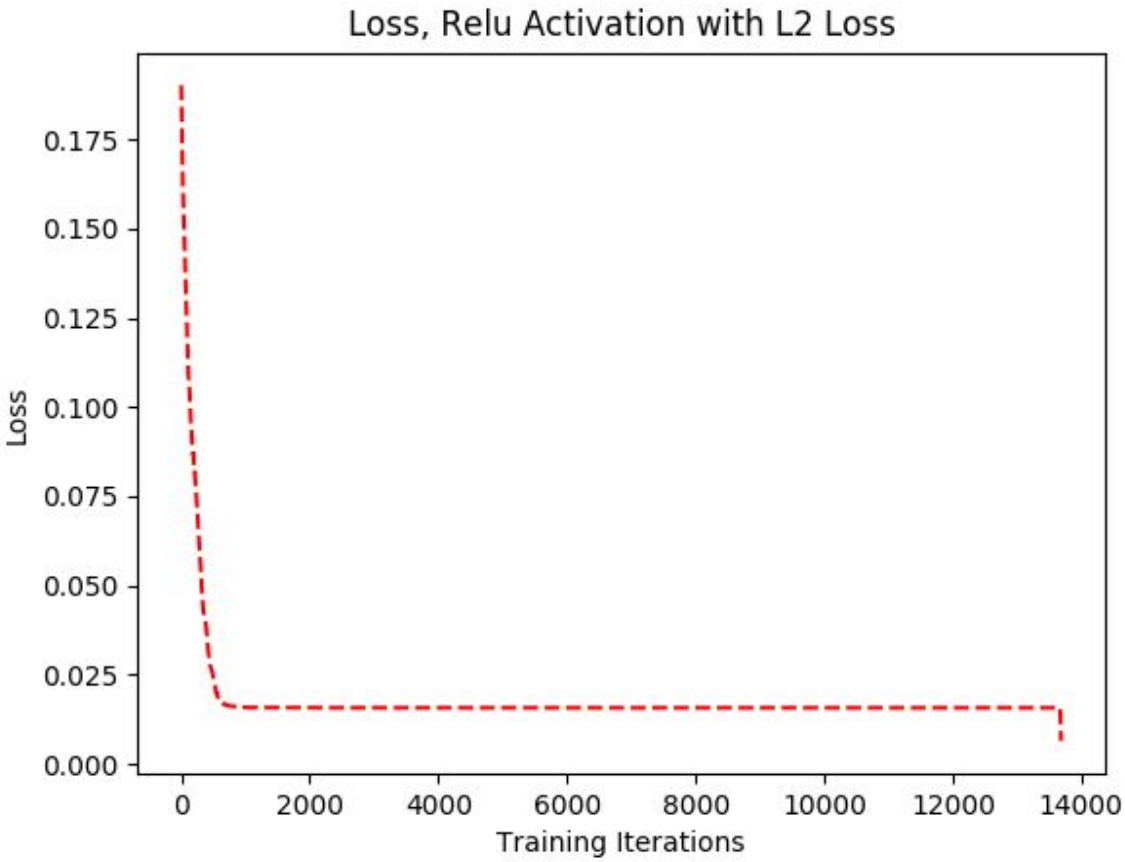

Accuracies, Sigmoid Activation with L2 Loss
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
Accuracy
0
1000
2000
3000
4000
Training Iterations

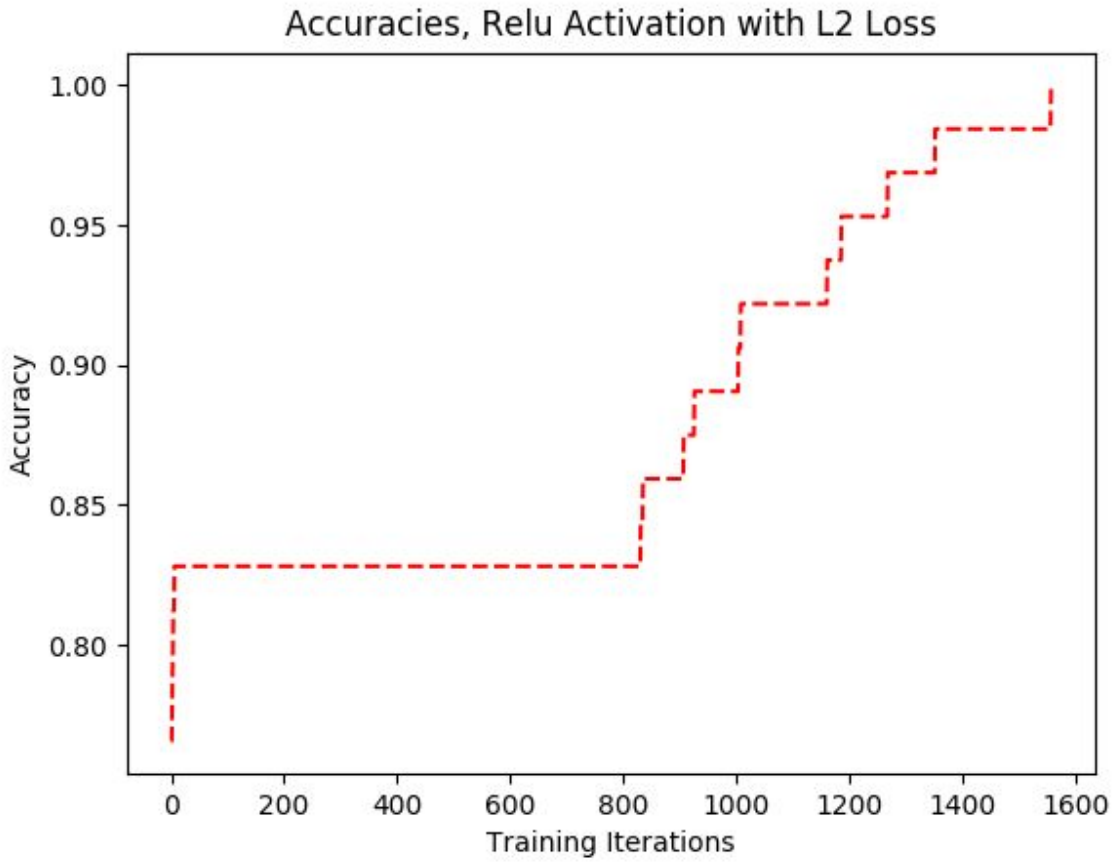
**P2.1.2:**


Loss, Sigmoid Activation with Binary Cross Entropy Loss
Loss
0.6
0.5
0.4
0.3
0.2
0.1
0
200
400
600
800
1000
1200
1400
1600
Training Iterations

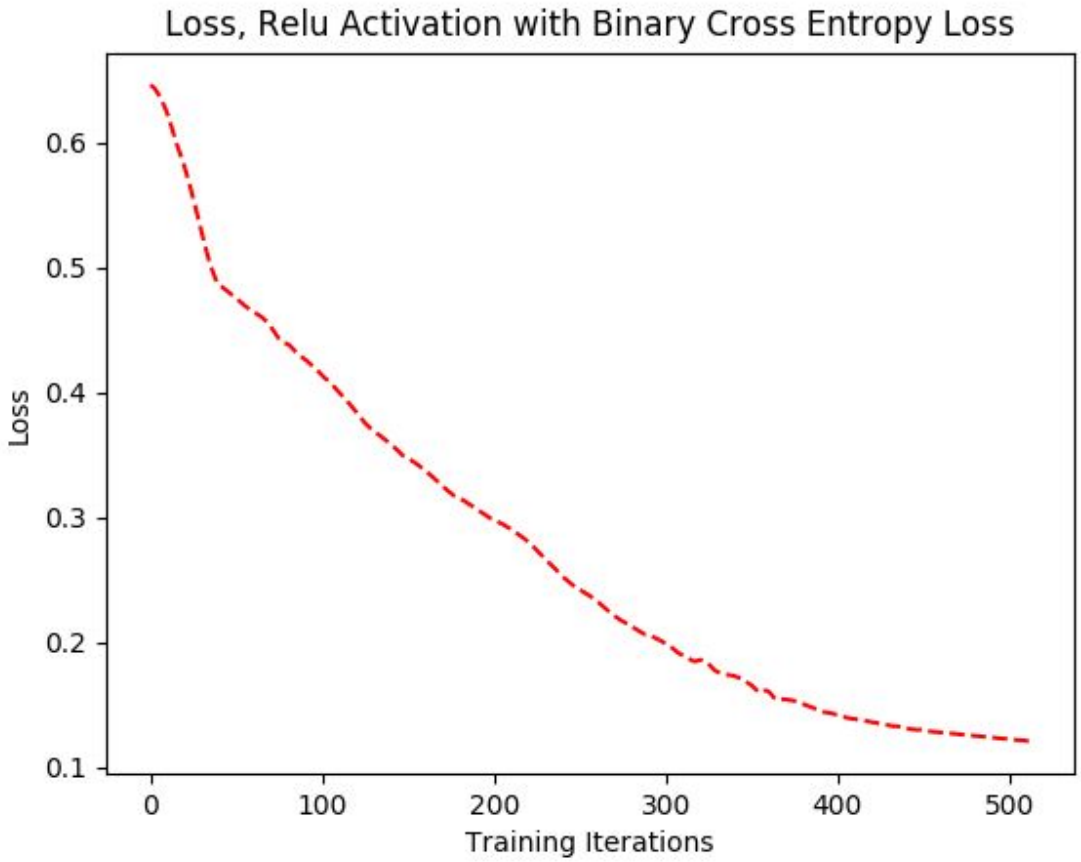

Accuracies, Sigmoid Activation with Binary Cross Entropy Loss
1.00
0.95
0.90
0.85
0.80
Accuracy
0
200
400
600
800
1000
1200
1400
1600
Training Iterations

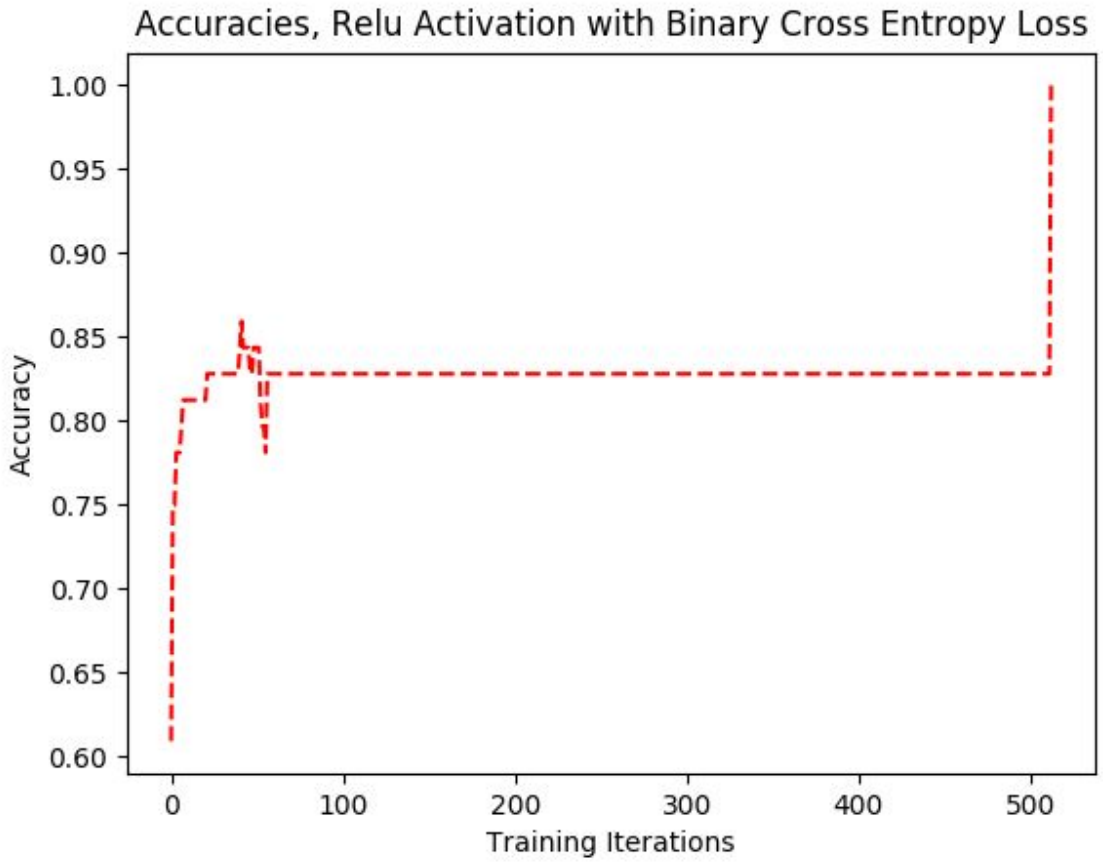
**P2.1.3:**


Loss, Relu Activation with L2 Loss
Loss
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
2000
4000
6000
8000
10000
12000
14000
Training Iterations


Accuracies, Relu Activation with L2 Loss
1.00
0.95
0.90
0.85
0.80
Accuracy
0
200
400
600
800
1000
1200
1400
1600
Training Iterations

**P2.1.4:**


Loss, Relu Activation with Binary Cross Entropy Loss
Loss
0.6
0.5
0.4
0.3
0.2
0.1
0
100
200
300
400
500
Training Iterations




**P2.1.5:**

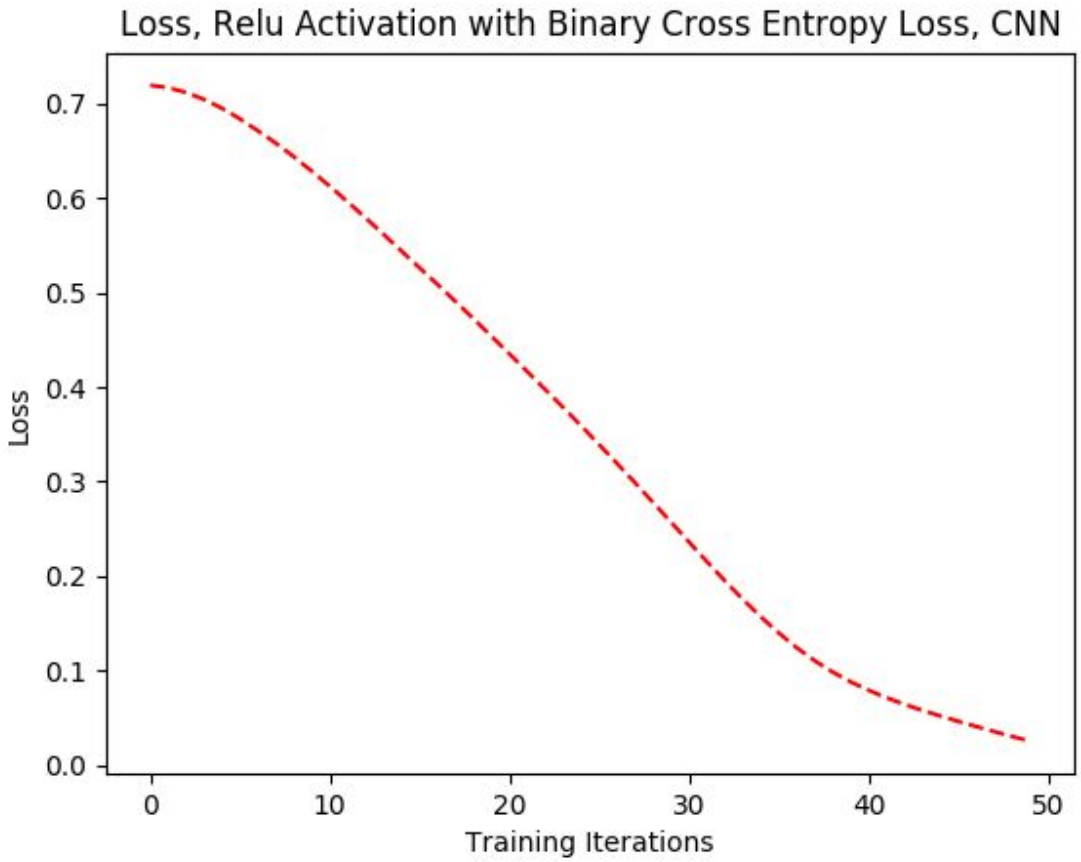
**The order is:**

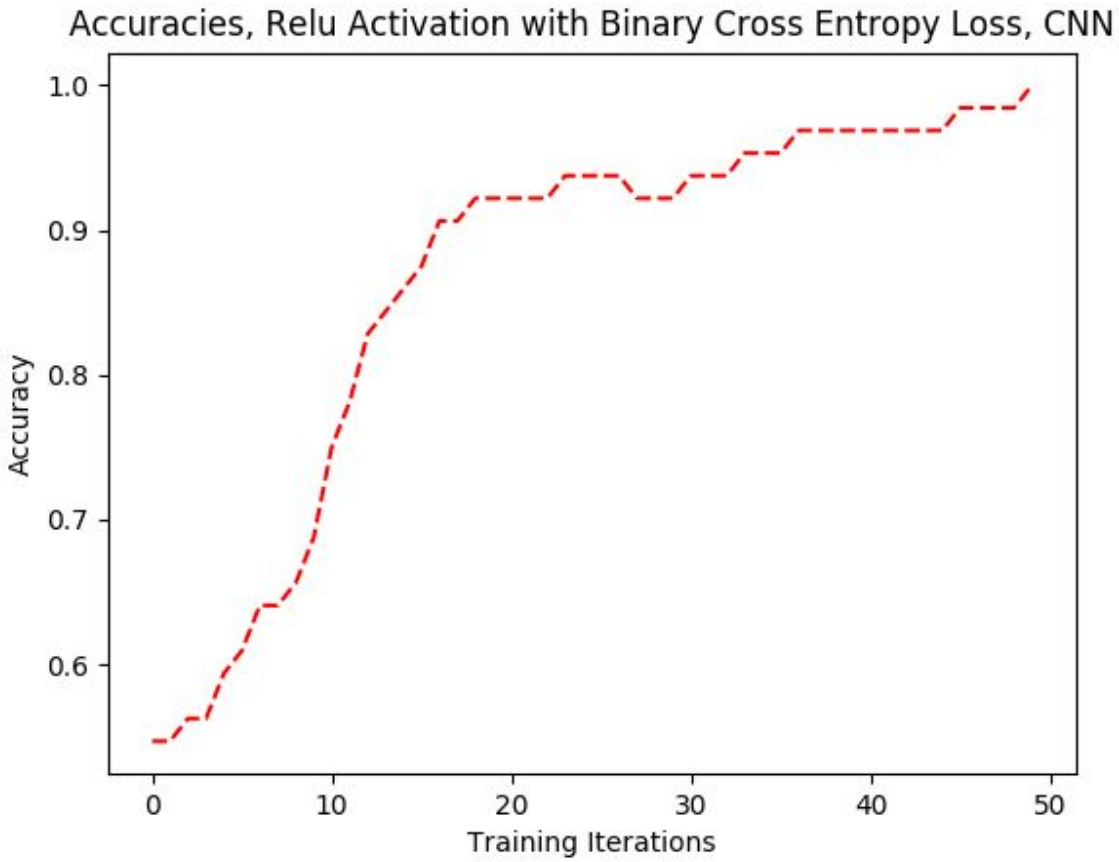
1. Relu with Cross Entropy Loss (50 iterations)
2. Sigmoid with Cross Entropy Loss (1558 iterations)
3. Sigmoid with L2 Loss (4378 iterations)
4. Relu with L2 Loss (13687 iterations)

**Explanation:**

For Cross Entropy, the loss falls very steepy because the further the predicted value is from the actual value, the higher the loss - the penalty is high. Relu converges faster than Sigmoid since it has non-saturation gradient, as opposed to Sigmoid's saturated gradient. Thus Relu with Cross-Entropy is the fastest to converge.

**P2.2:**


Loss, Relu Activation with Binary Cross Entropy Loss, CNN
Loss
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
10
20
30
40
50
Training Iterations


Accuracies, Relu Activation with Binary Cross Entropy Loss, CNN
Accuracy
1.0
0.9
0.8
0.7
0.6
0
10
20
30
40
50
Training Iterations