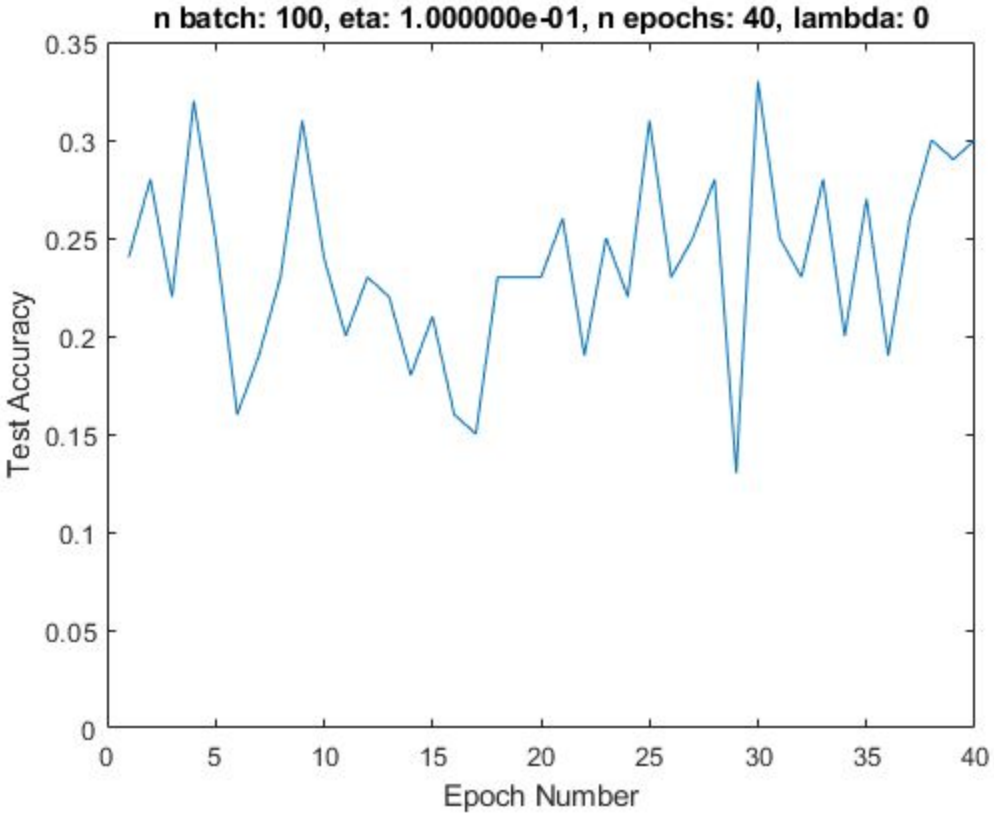
# Project 1 Report

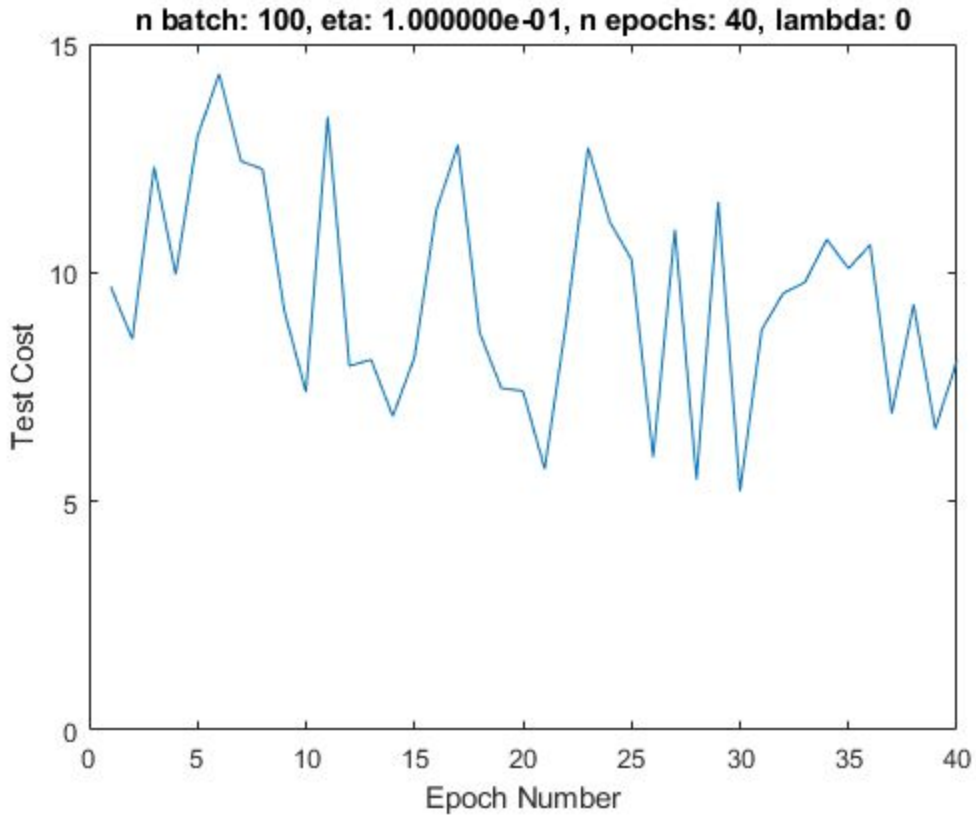
Jonas Wechsler, wechsler@kth.se

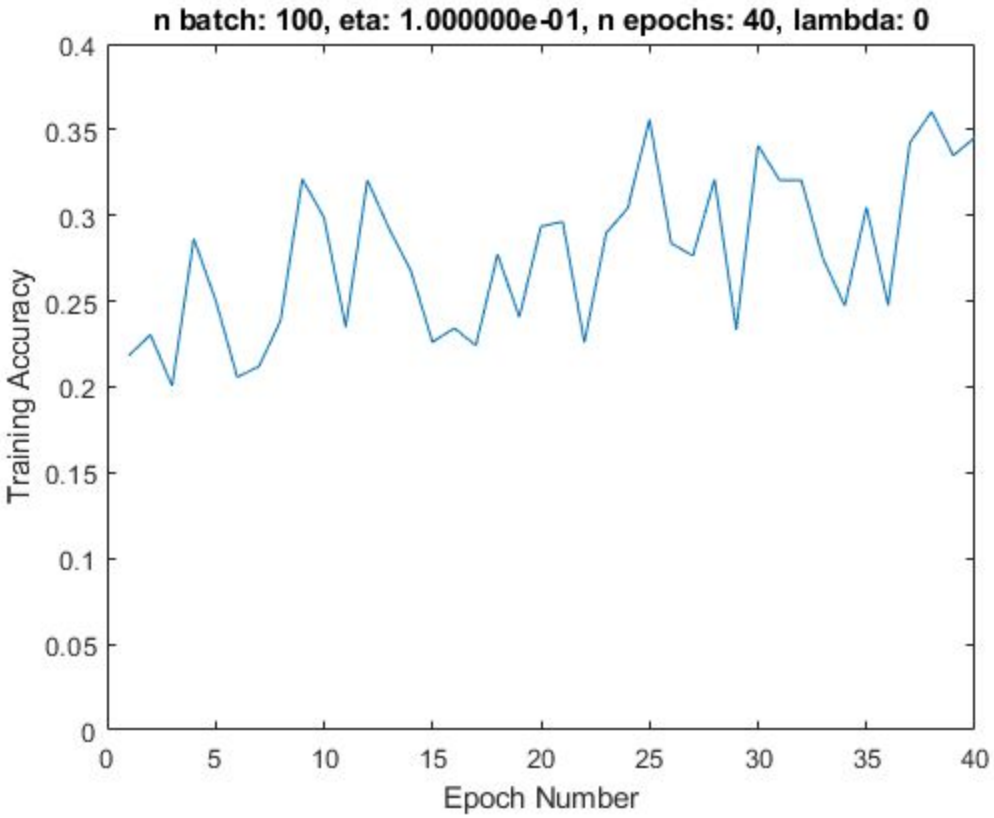
I successfully implemented the described functions. I observed the differences between the analytical and numerical gradients for batches of 1, 10, and 100, with varying subsets of the feature set, and confirmed that the max difference between the analytical and numerical data sets was less than 1e-6. Below are graphs and montages for Exercise 1. The hyperparameters are described in the titles of the graphs, and each set of graphs is preceded by its montage. Accuracy is measured from 0-1, so 0.4 is 40% accuracy. I found that increasing regularization reduced the accuracy of the model for this data set, with these training parameters. This is probably because this model is limited by its simplicity, so overtraining is less of an issue than undertraining. A learning rate that is too high (eta = 0.1) completely destroys the training process and makes the model never come close to converging.

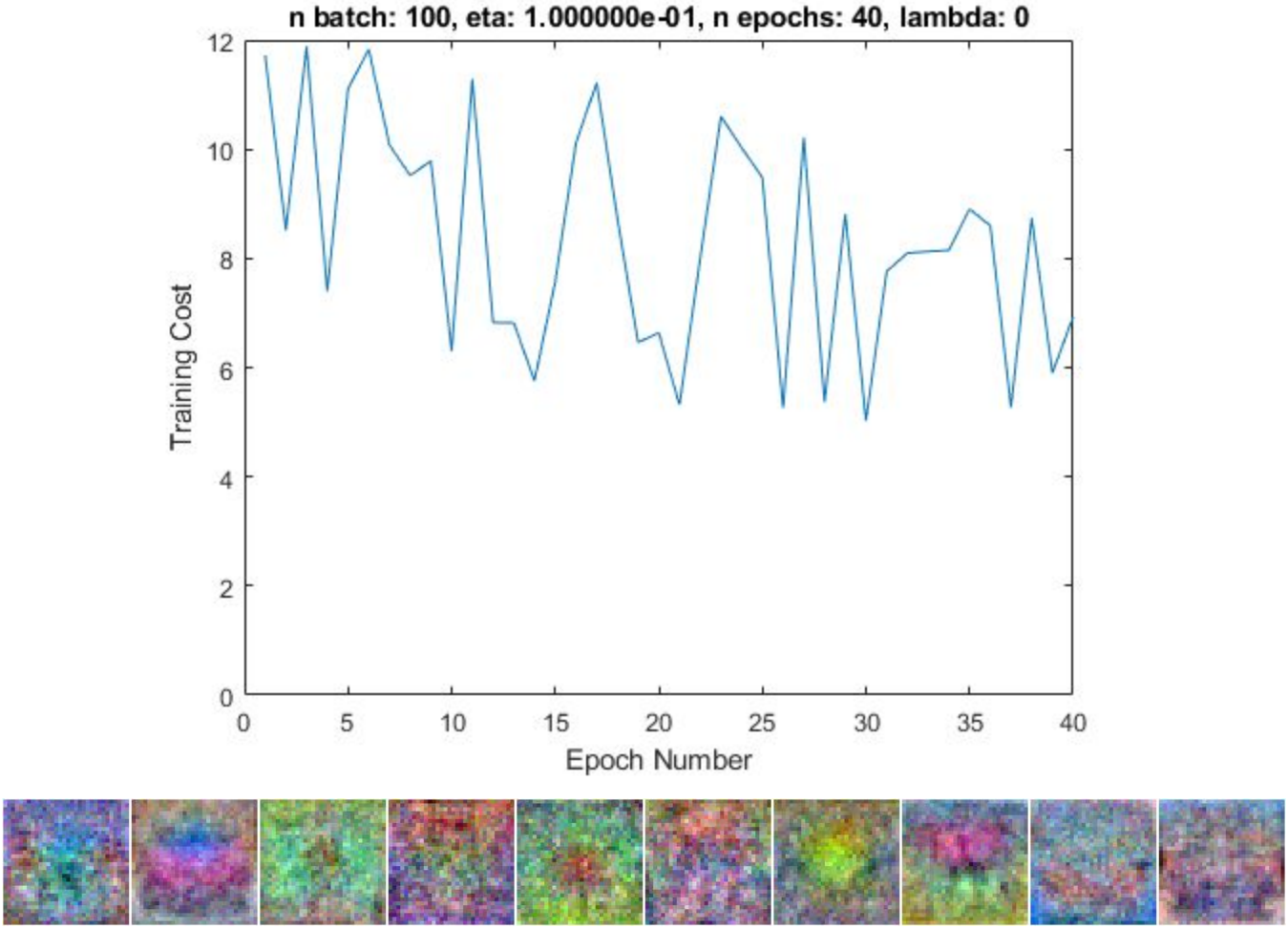
# Exercise 1 Data:

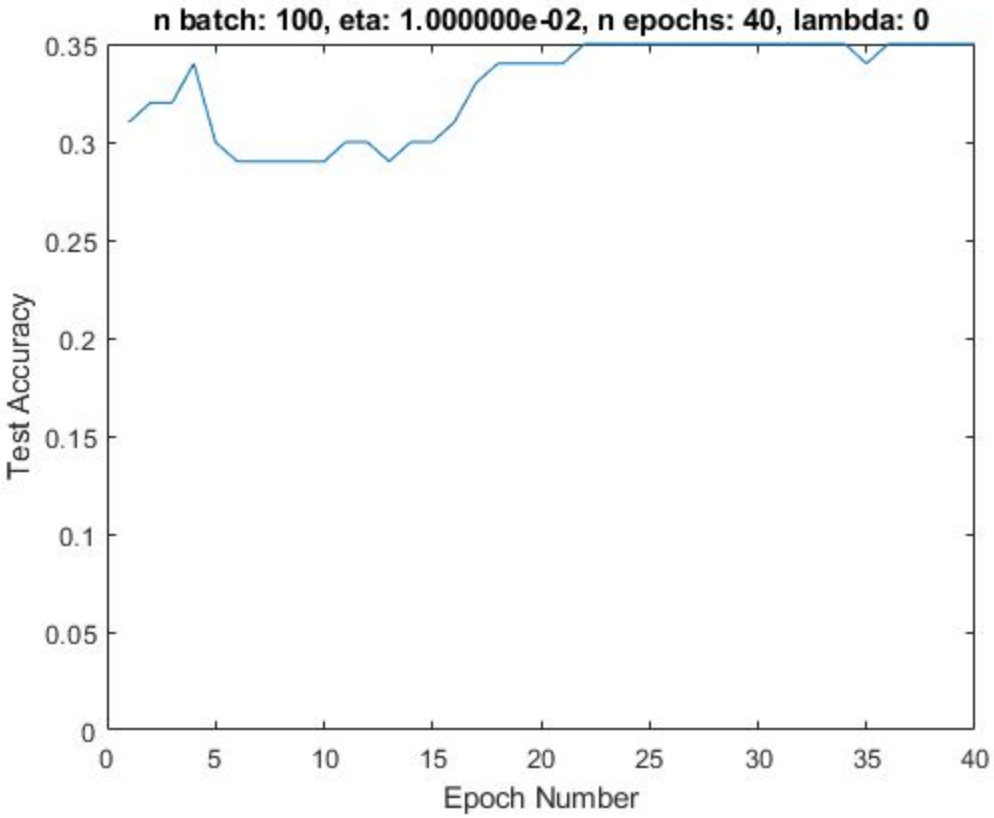
| n_batch | eta | n_epochs | lambda | Final accuracy | Final cost |
|---|---|---|---|---|---|
| 100 | 0.1 | 40 | 0 | 0.2460 | 7.0801 |
| 100 | 0.01 | 40 | 0 | 0.3630 | 2.0877 |
| 100 | 0.01 | 40 | 0.1 | 0.3380 | 2.0601 |
| 100 | 0.01 | 40 | 1 | 0.2580 | 2.1957 |

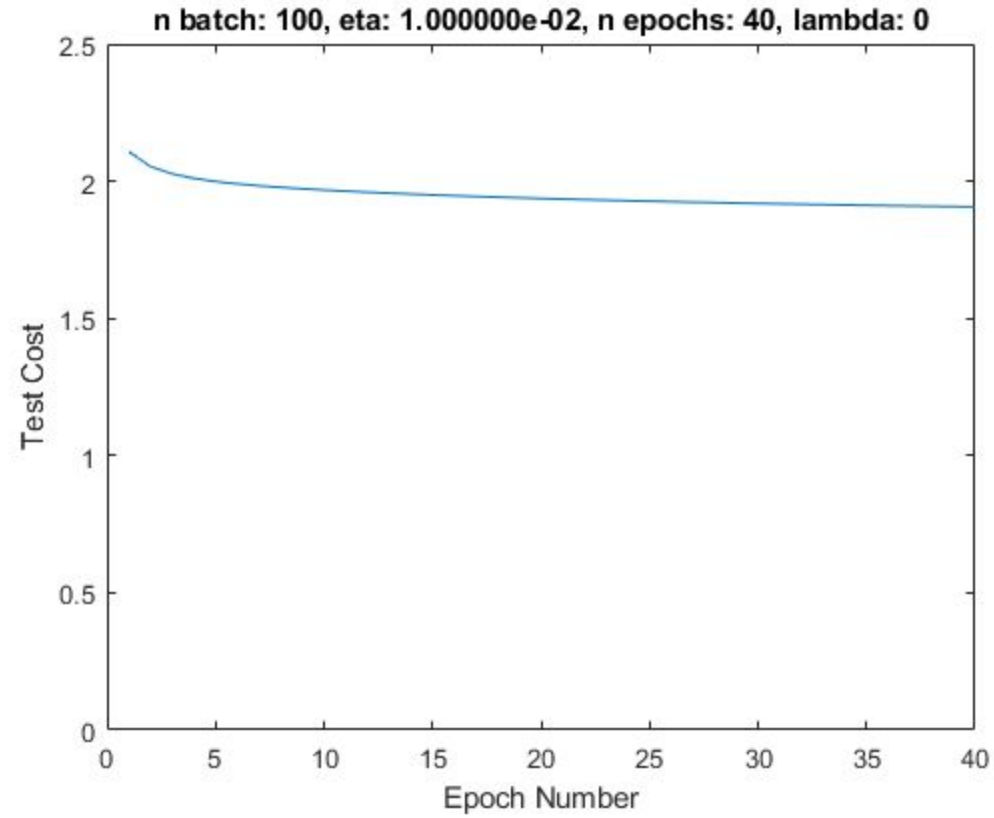

n batch: 100, eta: 1.000000e-01, n epochs: 40, lambda: 0
Test Accuracy
Epoch Number
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
30
35
40

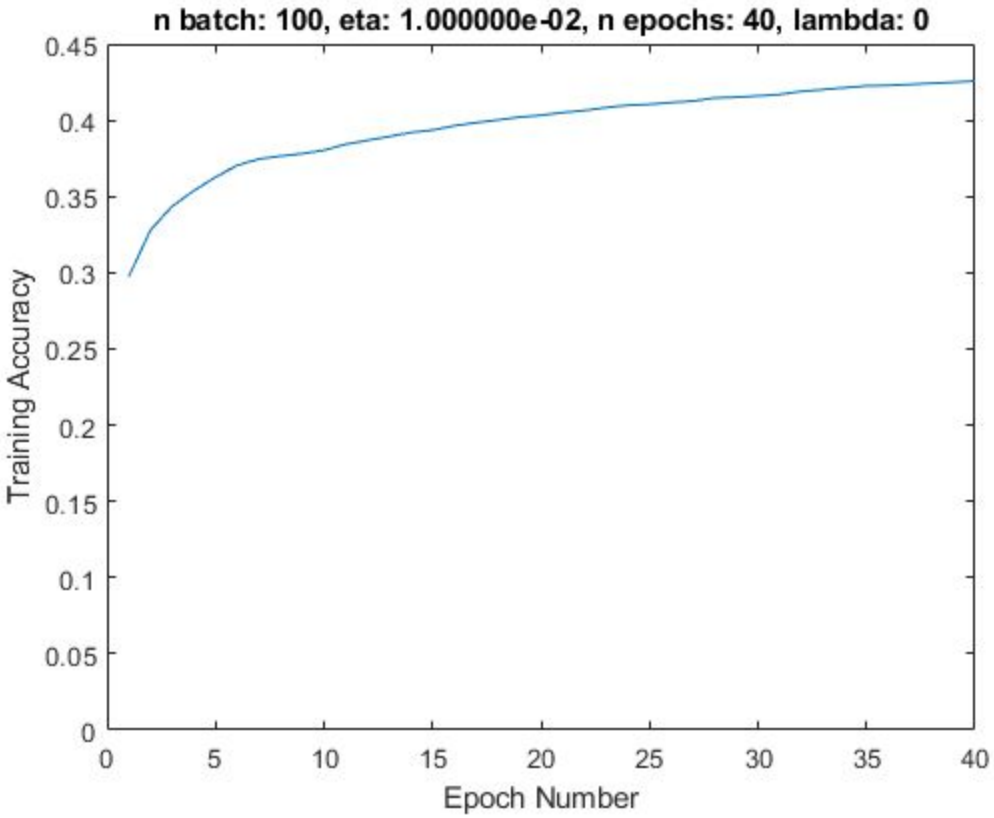

n batch: 100, eta: 1.000000e-01, n epochs: 40, lambda: 0
Test Cost
Epoch Number
0
5
10
15
0
5
10
15
20
25
30
35
40

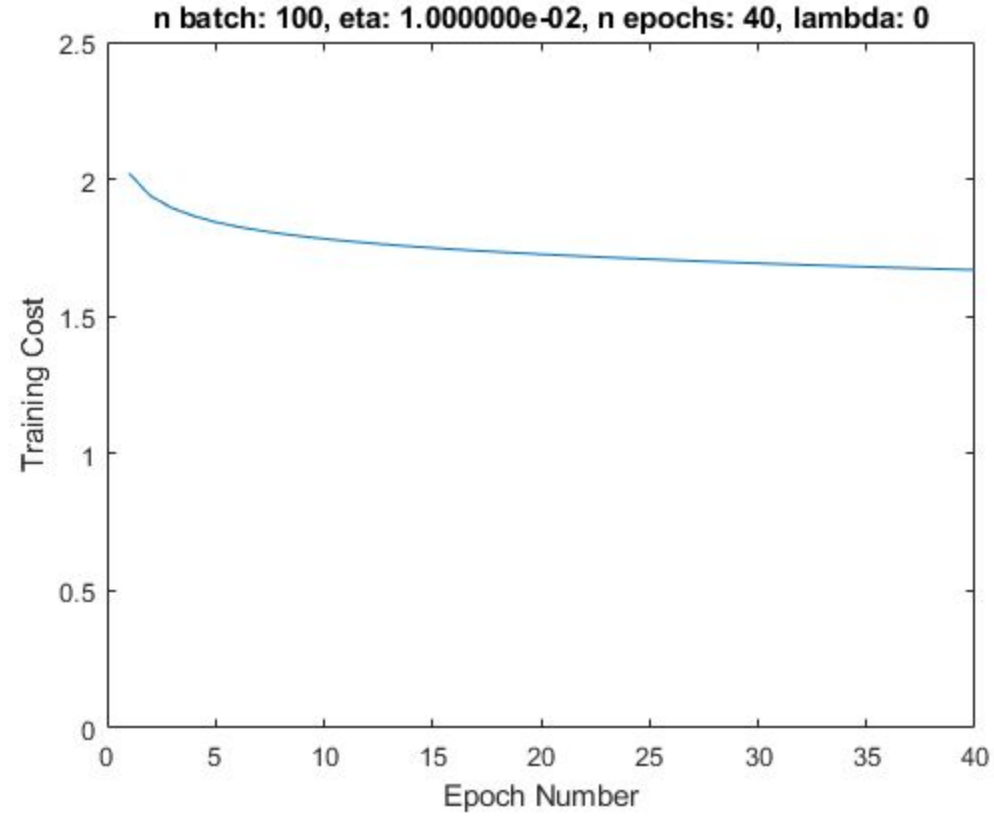

n batch: 100, eta: 1.000000e-01, n epochs: 40, lambda: 0
Training Accuracy
Epoch Number
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0
5
10
15
20
25
30
35
40

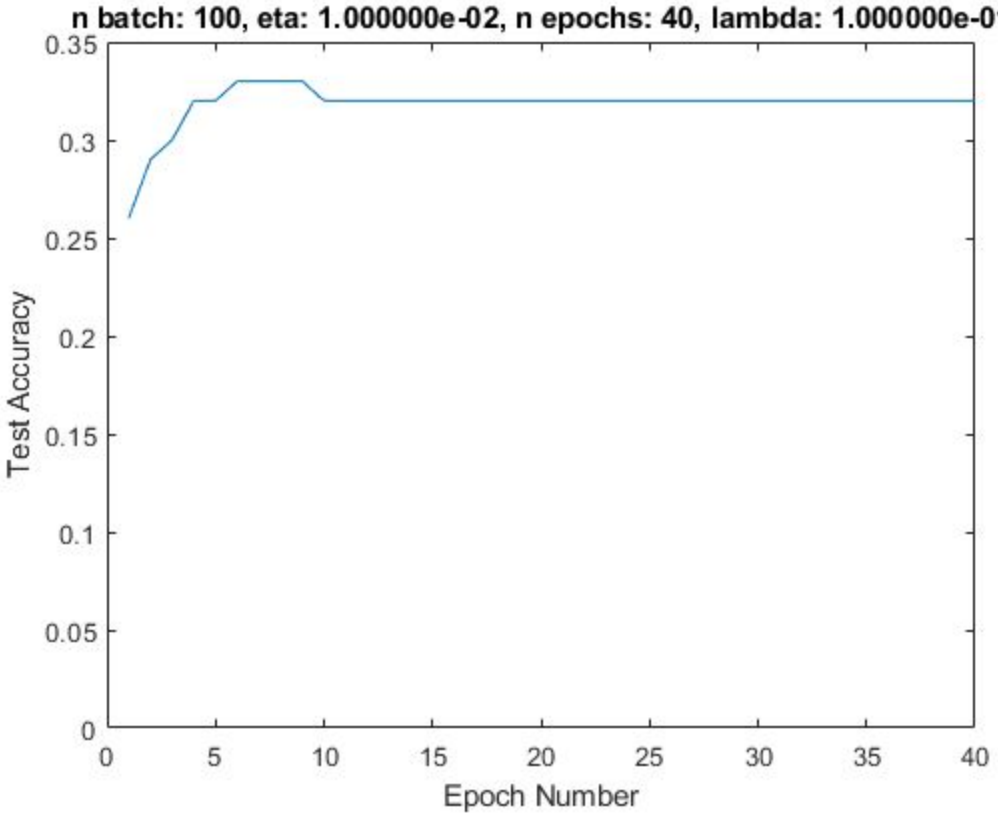

n batch: 100, eta: 1.000000e-01, n epochs: 40, lambda: 0
Training Cost
Epoch Number
0
2
4
6
8
10
12
0
5
10
15
20
25
30
35
40

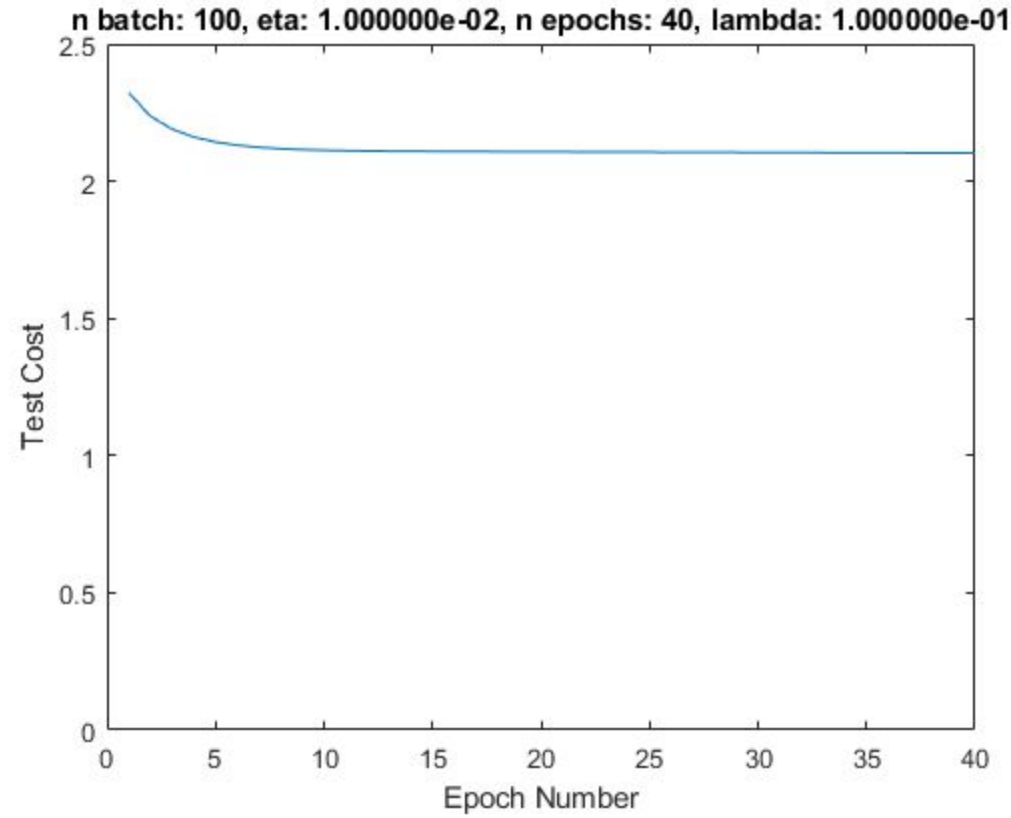

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 0
Test Accuracy
Epoch Number
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0
5
10
15
20
25
30
35
40

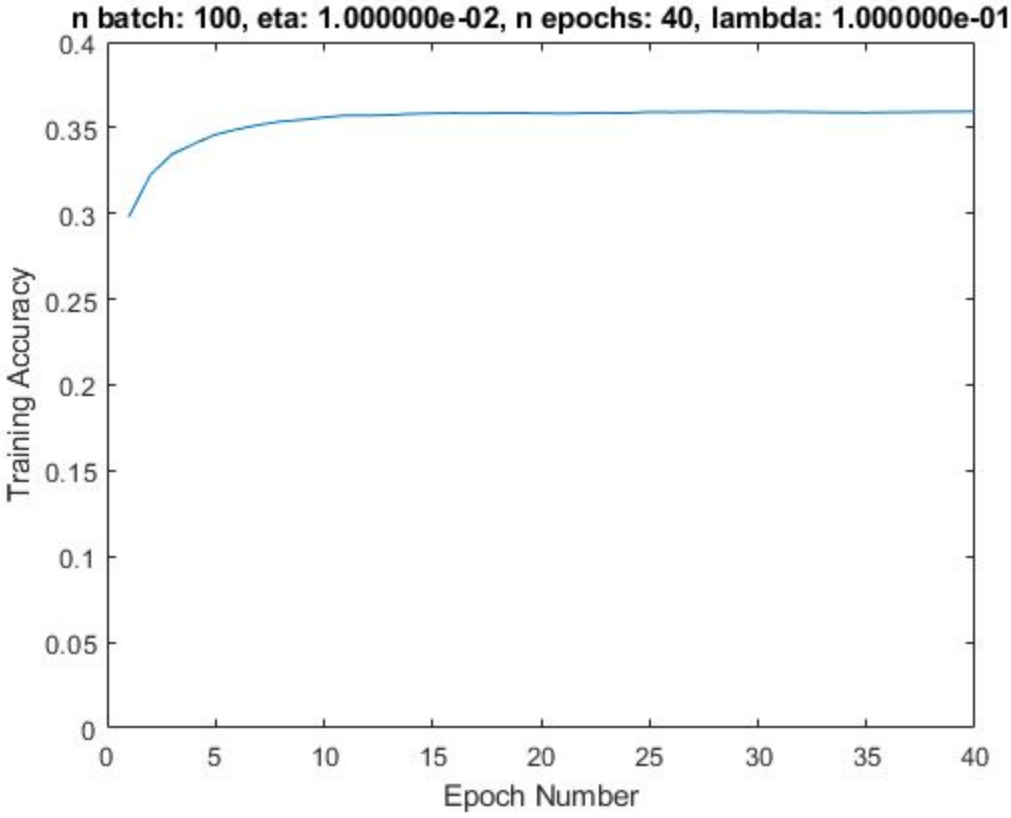

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 0
Test Cost
Epoch Number
0
0.5
1
1.5
2
2.5
0
5
10
15
20
25
30
35
40

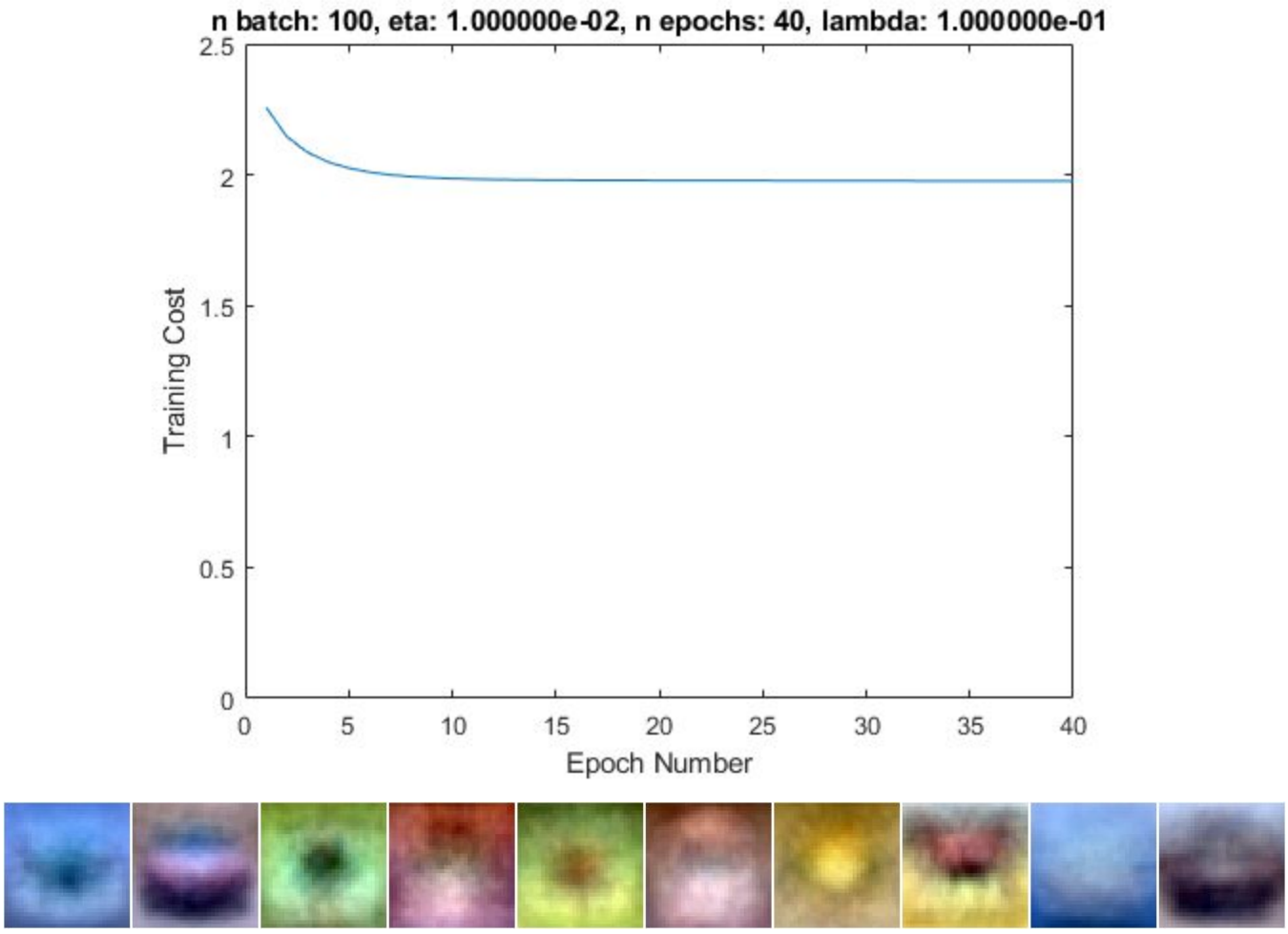

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 0
Training Accuracy
Epoch Number
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0
5
10
15
20
25
30
35
40

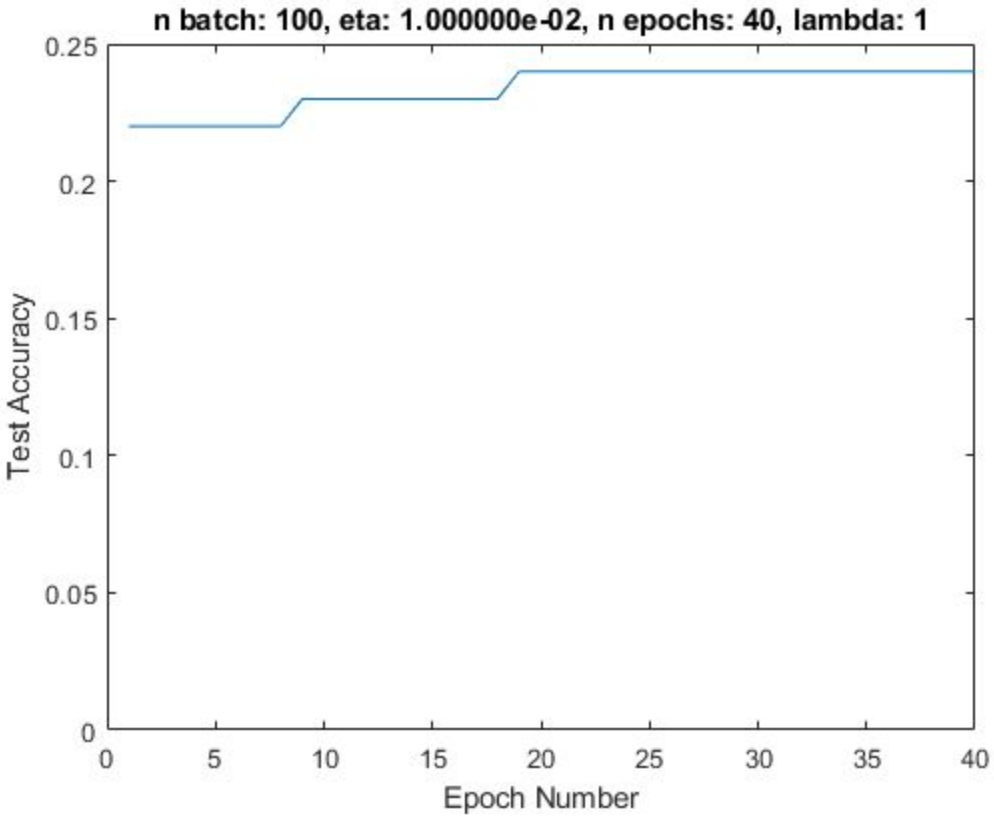

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 0
Training Cost
Epoch Number
0
0.5
1
1.5
2
2.5
0
5
10
15
20
25
30
35
40

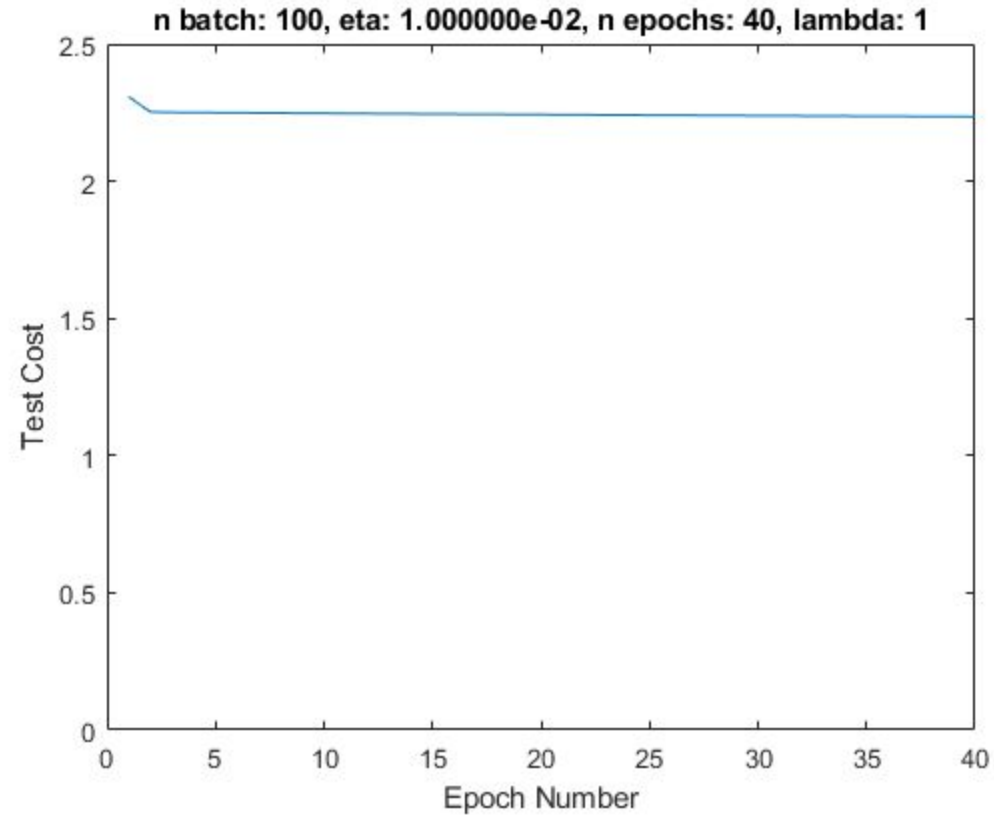

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 1.000000e-01
Test Accuracy
Epoch Number
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
30
35
40

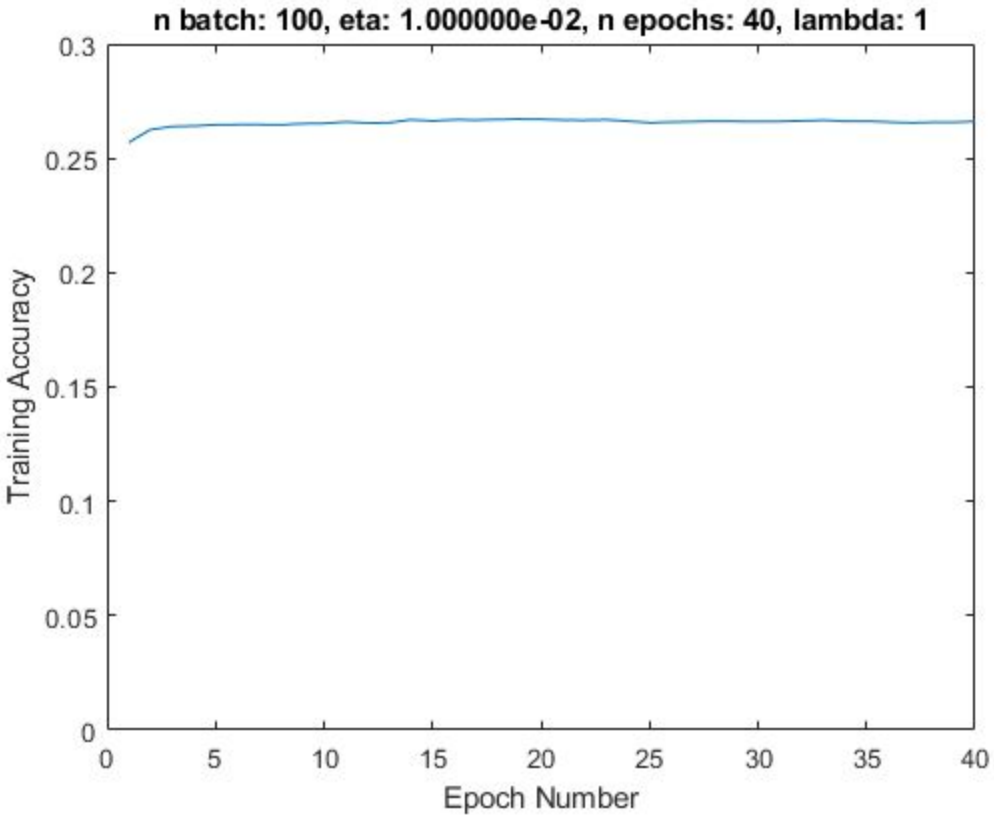

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 1.000000e-01
Test Cost
Epoch Number
0
0.5
1
1.5
2
2.5
0
5
10
15
20
25
30
35
40

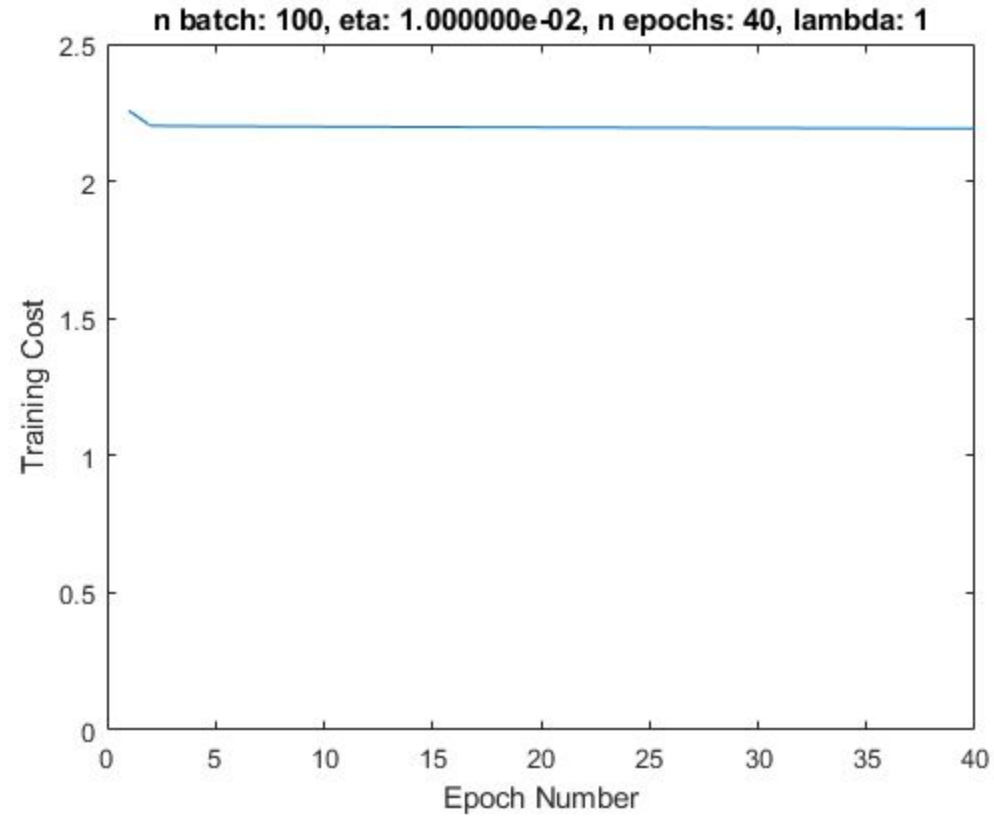

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 1.000000e-01
Training Accuracy
Epoch Number
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0
5
10
15
20
25
30
35
40


n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 1.000000e-01
Training Cost
Epoch Number
0
0.5
1
1.5
2
2.5
0
5
10
15
20
25
30
35
40


n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 1
Test Accuracy
Epoch Number
0
0.05
0.1
0.15
0.2
0.25
0
5
10
15
20
25
30
35
40

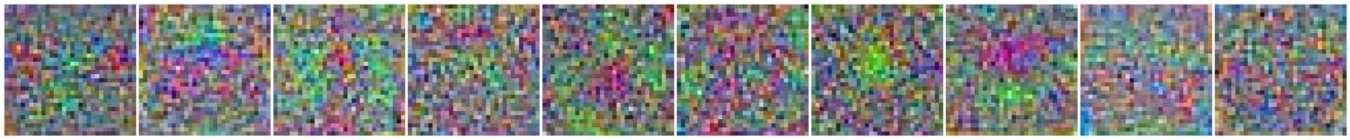

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 1
Test Cost
Epoch Number
0
0.5
1
1.5
2
2.5
0
5
10
15
20
25
30
35
40

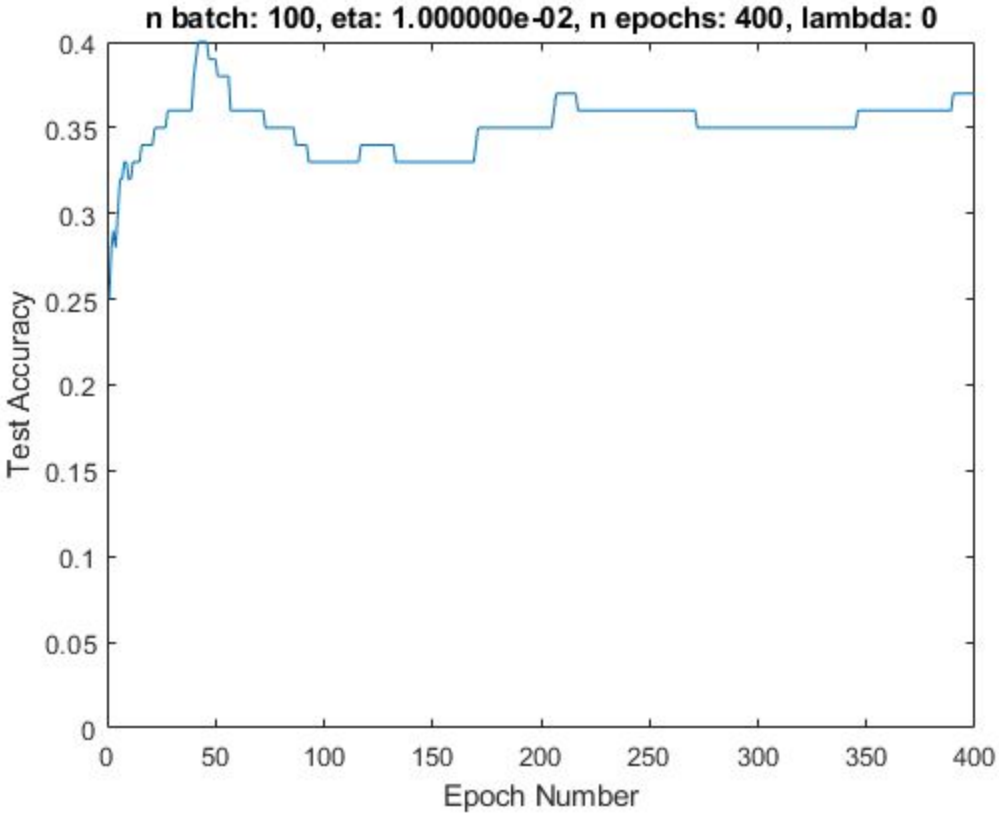

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 1
Training Accuracy
Epoch Number
0
0.05
0.1
0.15
0.2
0.25
0.3
0
5
10
15
20
25
30
35
40

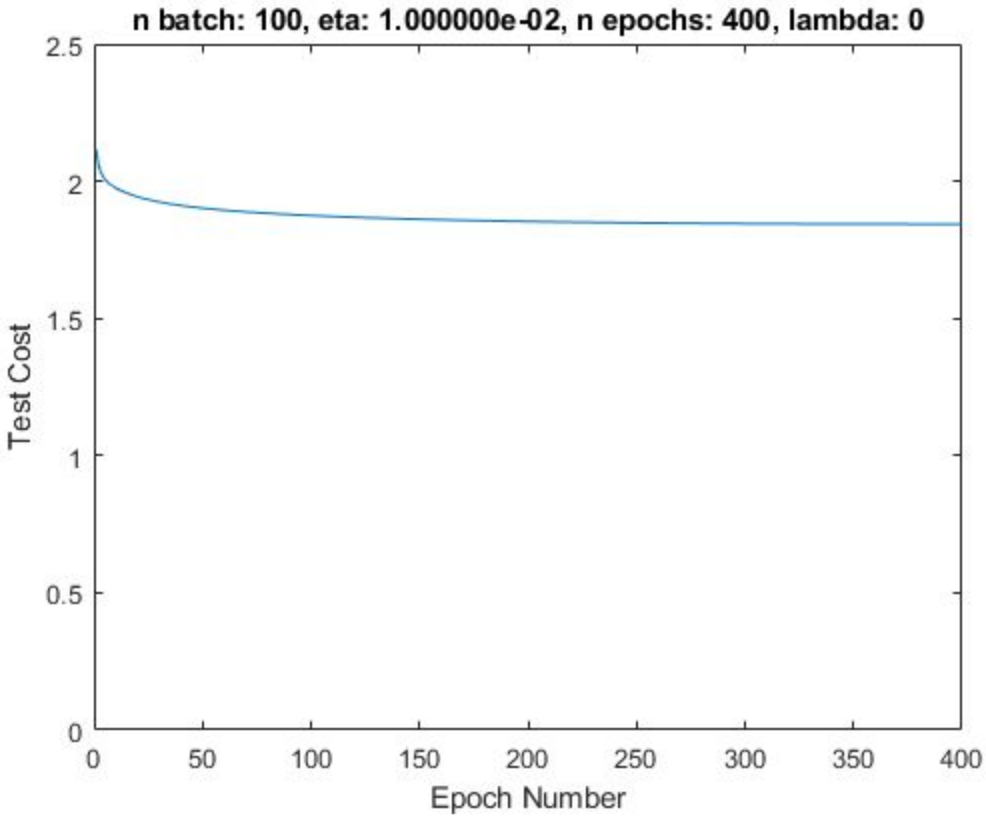

n batch: 100, eta: 1.000000e-02, n epochs: 40, lambda: 1
Training Cost
Epoch Number
0
0.5
1
1.5
2
2.5
0
5
10
15
20
25
30
35
40

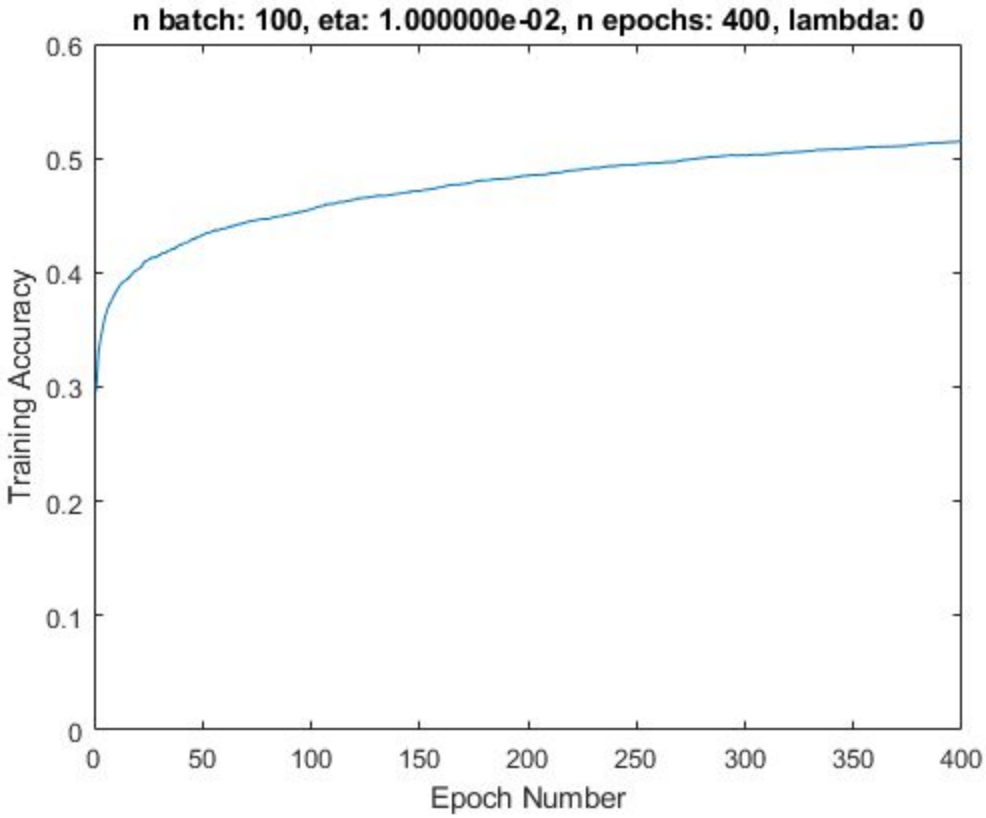
# Exercise 2

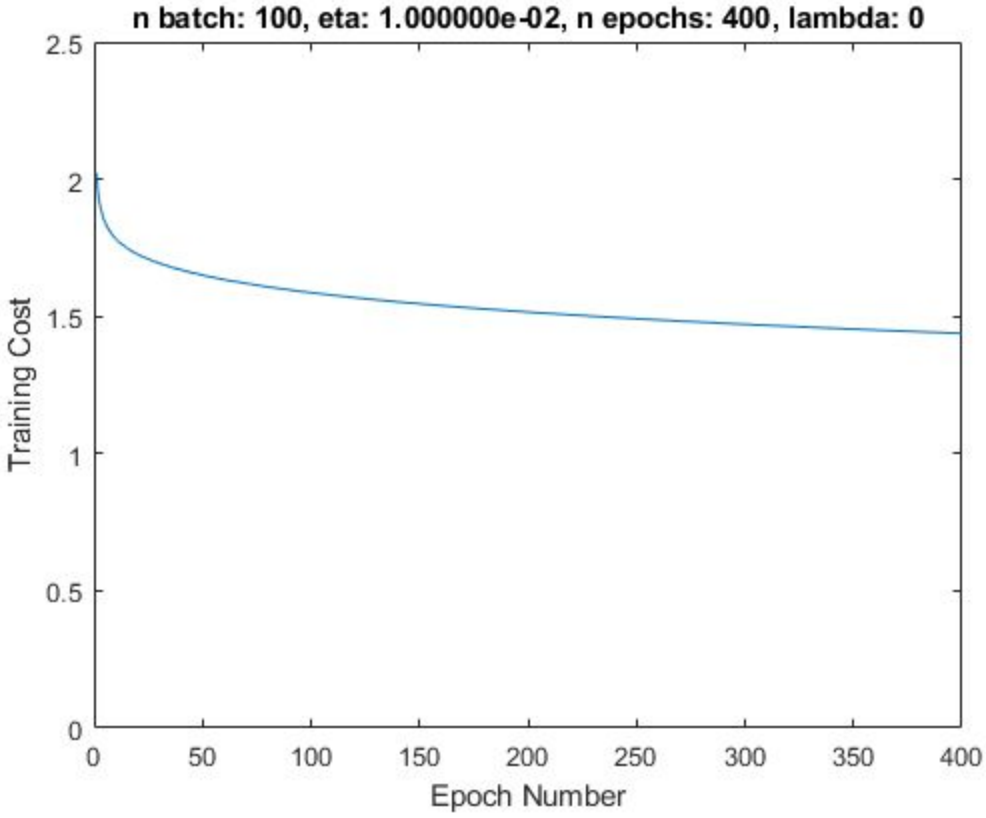
In exercise 1, without any optimizations the best accuracy 0.36. For exercise 2, I first ran the trainer with eta decaying by a factor of 0.9 for 40 epochs, which yielded a test accuracy of 0.3750. I then ran the trainer with each data set, which gave a 0.4070 accuracy. I then ran the data for 400 epochs, keeping track of the most accurate weights and using them to calculate the final accuracy. The final accuracy was 0.3950. The graphs for this 400 epoch run are below, followed by graphs for a 1000 epoch run. The 1000 epoch run has eta decay of 0.99, and the best run was chosen to calculate accuracy. Its accuracy was 0.4010. I then trained on all 5 data sets, used an eta decay rate of 0.99, and trained for 200 epochs, keeping track of the best weights. This yielded an accuracy of 0.4210. The best accuracy was obtained by the last model, which yielded an accuracy of **42.1%.**

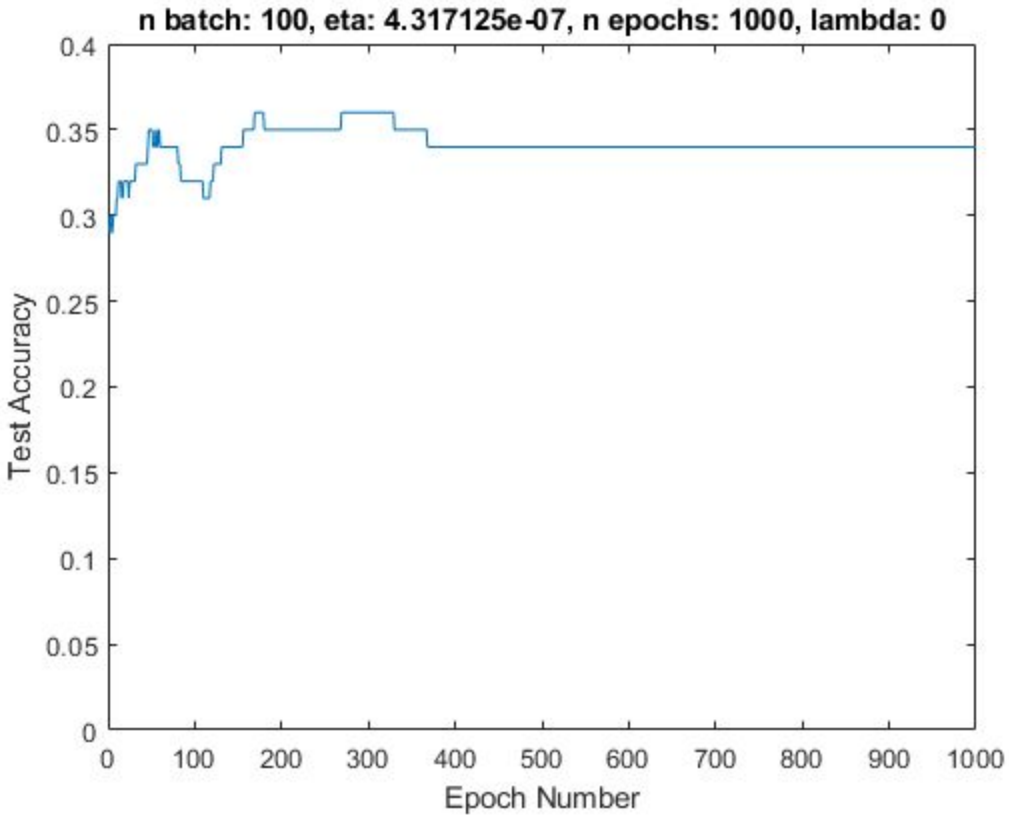


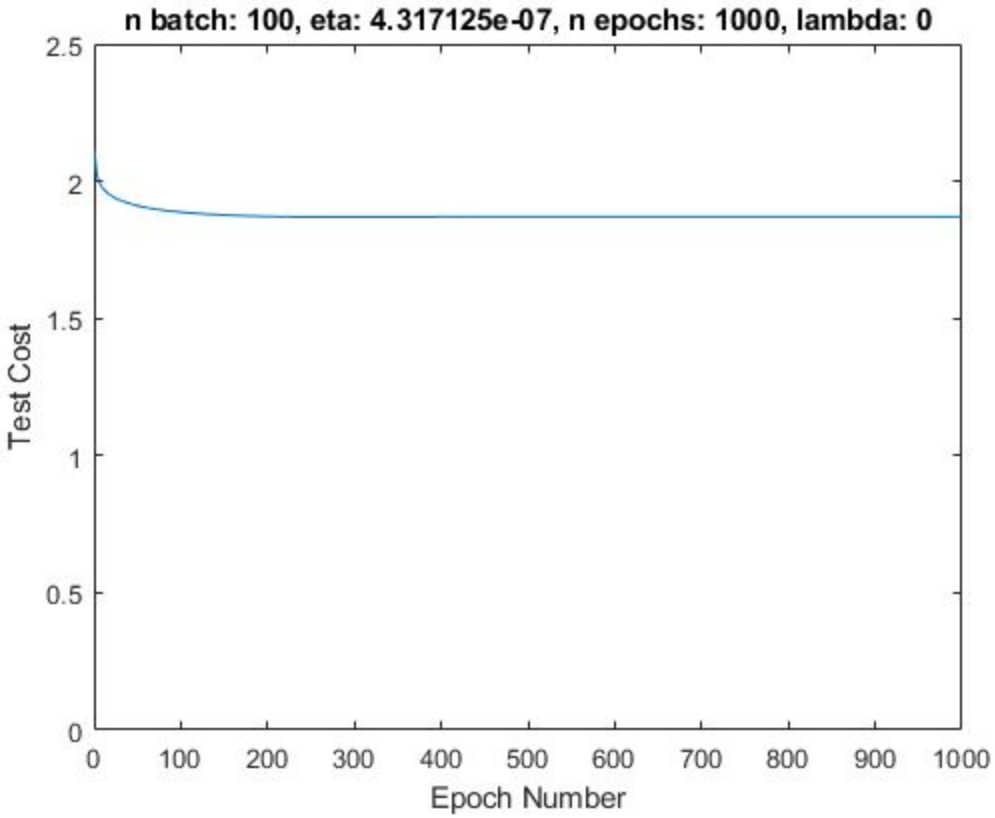


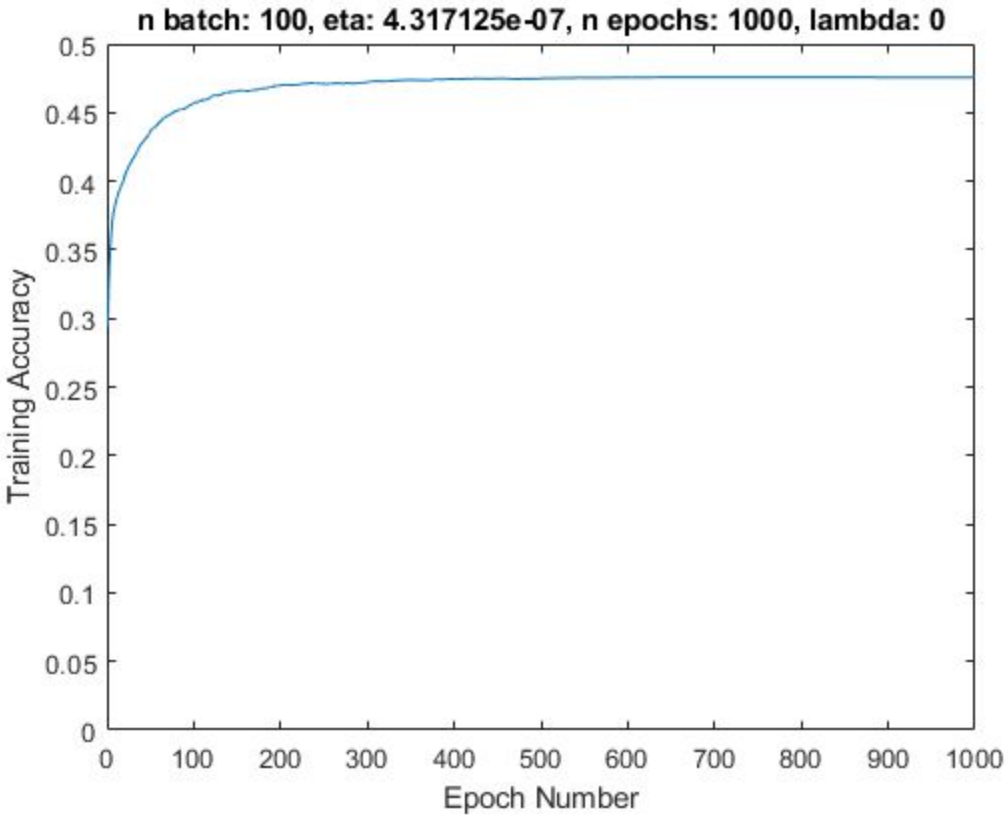

n batch: 100, eta: 1.000000e-02, n epochs: 400, lambda: 0
Test Cost
Epoch Number
2.5
2
1.5
1
0.5
0
0
50
100
150
200
250
300
350
400

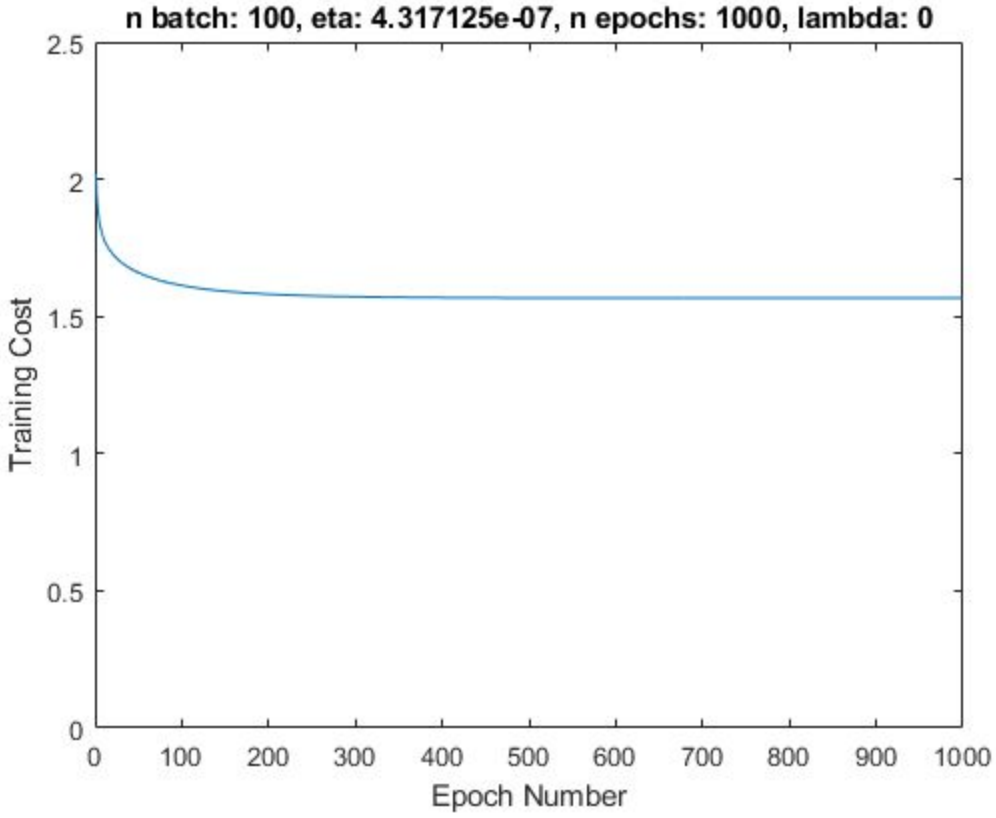

n batch: 100, eta: 1.000000e-02, n epochs: 400, lambda: 0
Training Accuracy
Epoch Number
0
0.1
0.2
0.3
0.4
0.5
0.6
0
50
100
150
200
250
300
350
400


n batch: 100, eta: 1.000000e-02, n epochs: 400, lambda: 0
Training Cost
Epoch Number
0
0.5
1
1.5
2
2.5
0
50
100
150
200
250
300
350
400


n batch: 100, eta: 4.317125e-07, n epochs: 1000, lambda: 0
Test Accuracy
Epoch Number
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0
100
200
300
400
500
600
700
800
900
1000


n batch: 100, eta: 4.317125e-07, n epochs: 1000, lambda: 0
Test Cost
Epoch Number
2.5
2
1.5
1
0.5
0
0
100
200
300
400
500
600
700
800
900
1000


n batch: 100, eta: 4.317125e-07, n epochs: 1000, lambda: 0
Training Accuracy
Epoch Number
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0
100
200
300
400
500
600
700
800
900
1000


n batch: 100, eta: 4.317125e-07, n epochs: 1000, lambda: 0
Training Cost
Epoch Number
0
0.5
1
1.5
2
2.5
0
100
200
300
400
500
600
700
800
900
1000