CS615 Homework 3

Indervir Singh (is386)

May 2020

1. Theory Answers

1. (a) The first step is to flip K horizontally, then vertically: $\begin{bmatrix} 9 & 8 & 7 \\ 6 & 4 & 5 \\ 3 & 2 & 1 \end{bmatrix}$

   (b) Then we have to multiple X and the flipped K: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} * \begin{bmatrix} 9 & 8 & 7 \\ 6 & 4 & 5 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 16 & 21 \\ 24 & 25 & 24 \\ 21 & 16 & 9 \end{bmatrix}$

   (c) Then we sum the resulting matrix, and get the feature map: $\begin{bmatrix} 165 \end{bmatrix}$

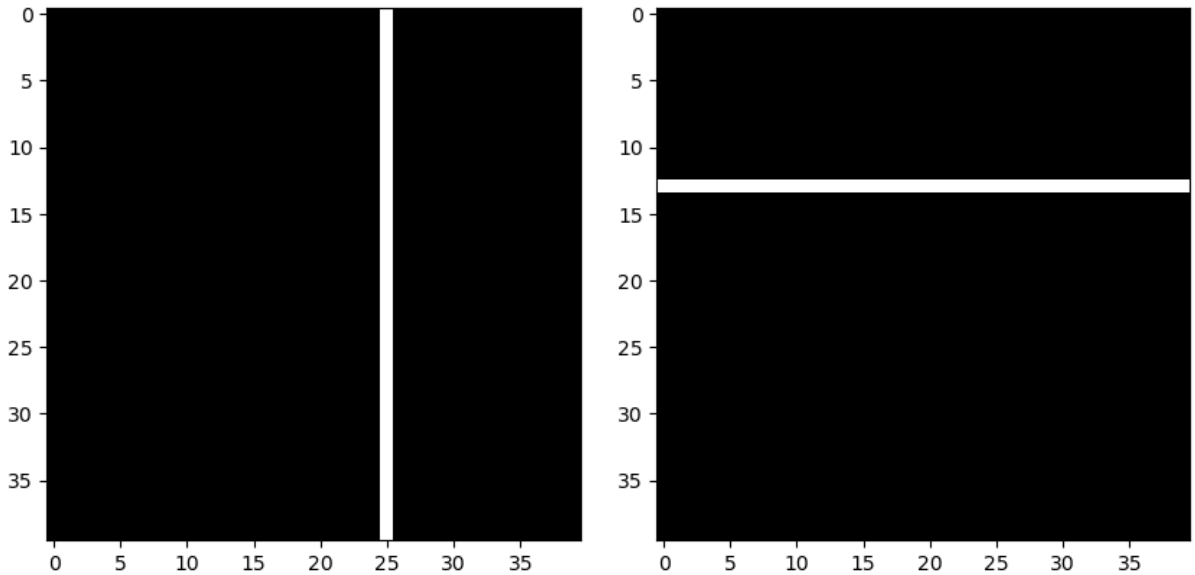
2. (a) Max Pooling = $\begin{bmatrix} 5 & 6 & 12 \\ 8 & 9 & 4 \end{bmatrix}$

   (b) Mean Pooling = $\begin{bmatrix} 3.00 & 3.25 & 3.75 \\ -46.25 & -47.50 & -49.00 \end{bmatrix}$

3. The only kernel that would result in X * K = X, would be K = $\begin{bmatrix} 1 \end{bmatrix}$. This is because the multiplication would result in the same numeric values, and since the width/stride of the convolution are both 1, the image does not get "summarized" into a smaller matrix.

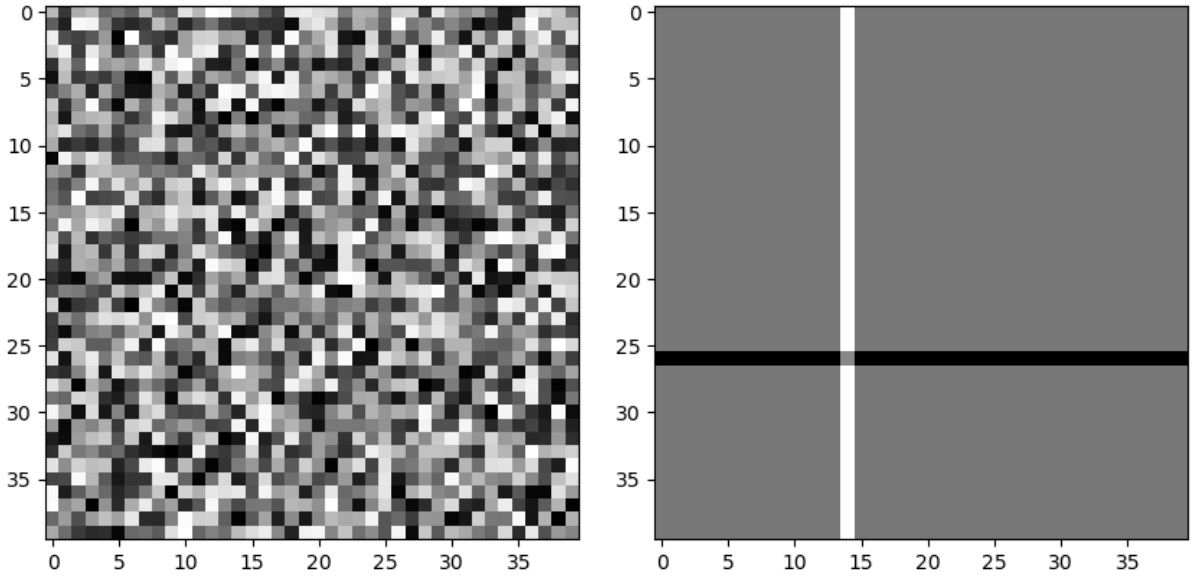
Synthetic Data

This is the synthetic data I created for parts 2-5.

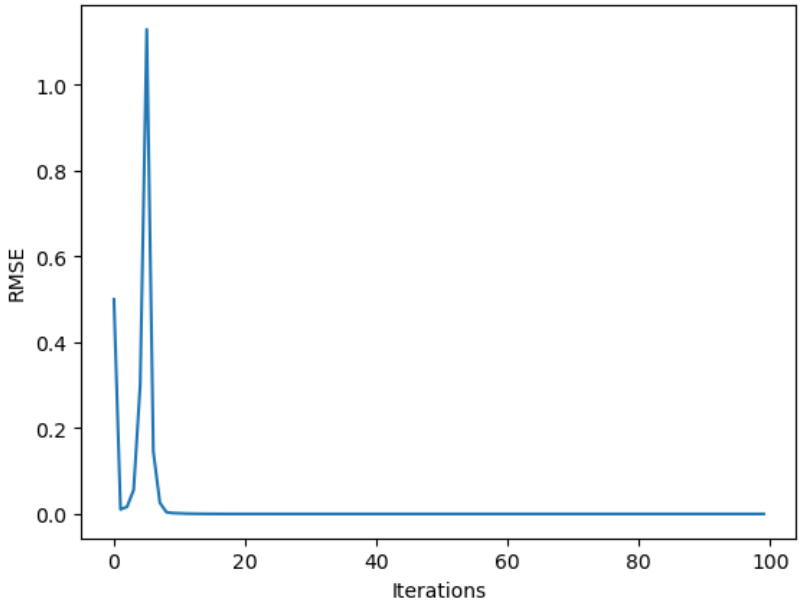


2. CNN for LSE Classification

1. Image representations of the initial and final kernels.



2. Plot of the RMSE as a function of the number of iterations.



3. Hyper Parameters

   (a) Learning Rate = 0.1

   (b) Termination Criteria = 100 iterations

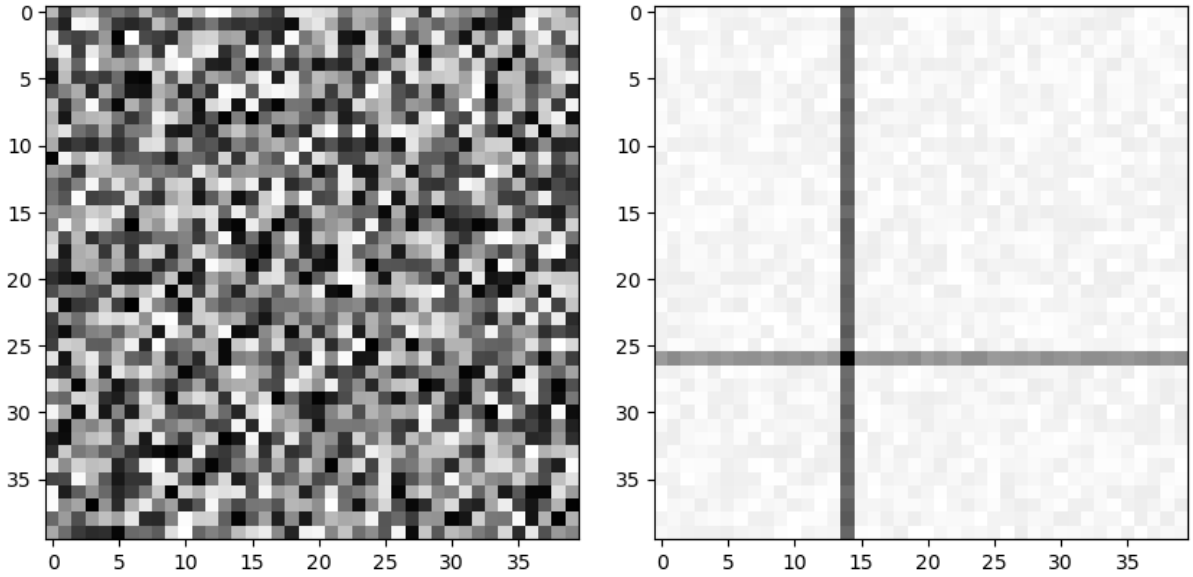
   (c) Regularization Term = 0.0001

   (d) Kernel was initialized to values between -0.00001 and 0.00001.

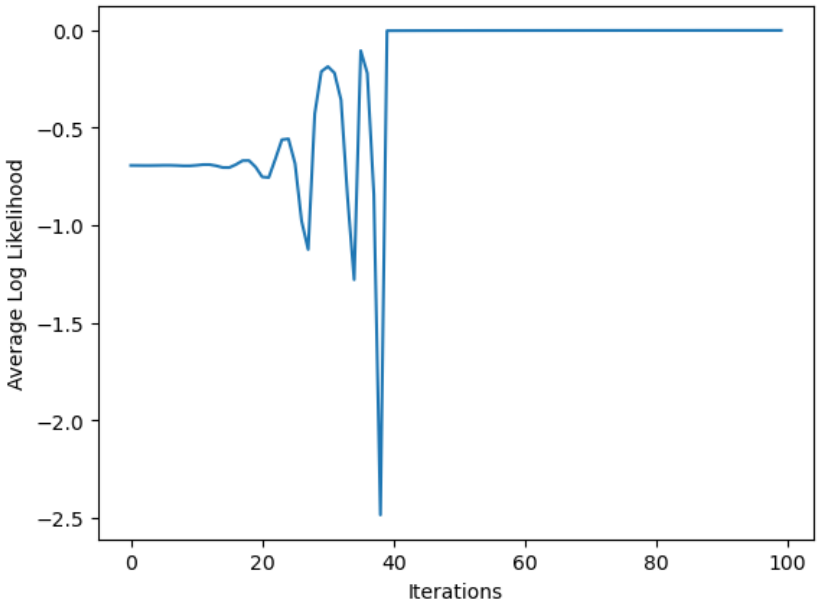
   (e) Theta was initialized to values between -1 and 1.

3. CNN for MLE Classification

1. Image representations of the initial and final kernels.



2. Plot of the Average Log Likelihood as a function of the number of iterations.



3. Hyper Parameters

   (a) Learning Rate = 0.001

   (b) Termination Criteria = 100 iterations

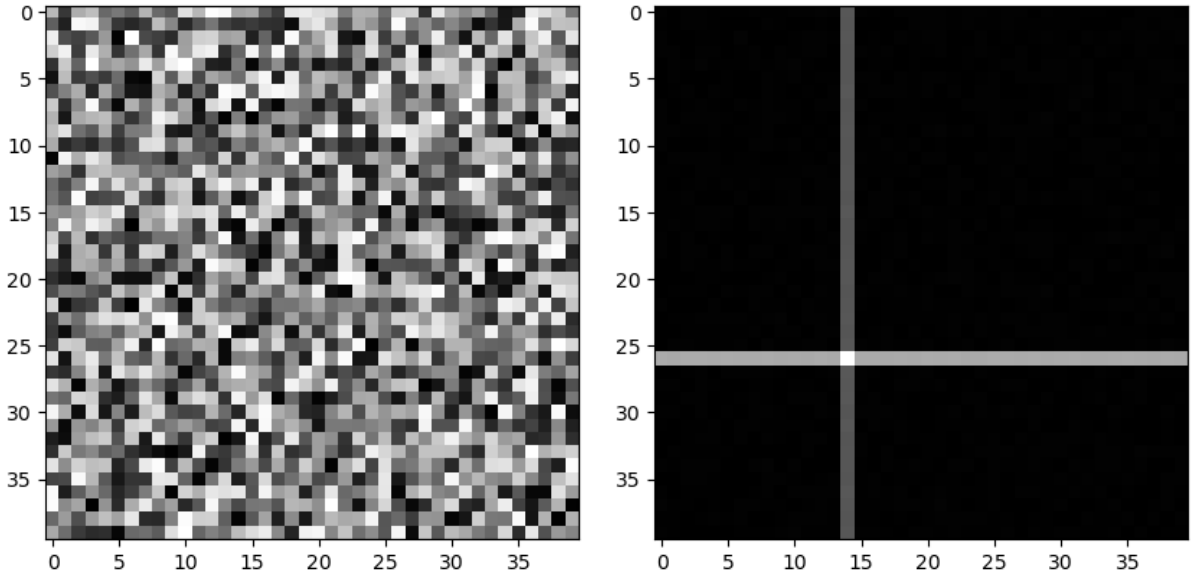
   (c) Regularization Term = 0.1

   (d) Kernel was initialized to values between -0.001 and 0.001.

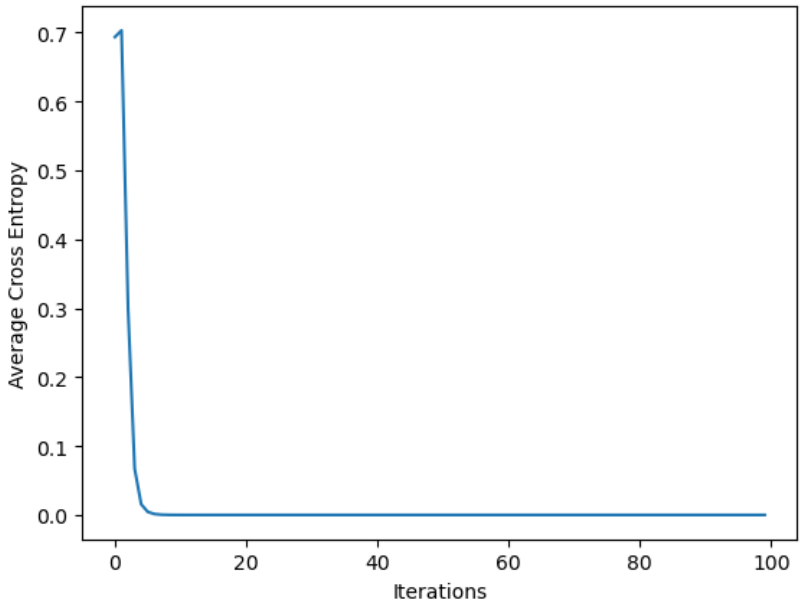
   (e) Theta was initialized to values between -0.001 and 0.001.

4. CNN for LCE Classification

1. Image representations of the initial and final kernels.



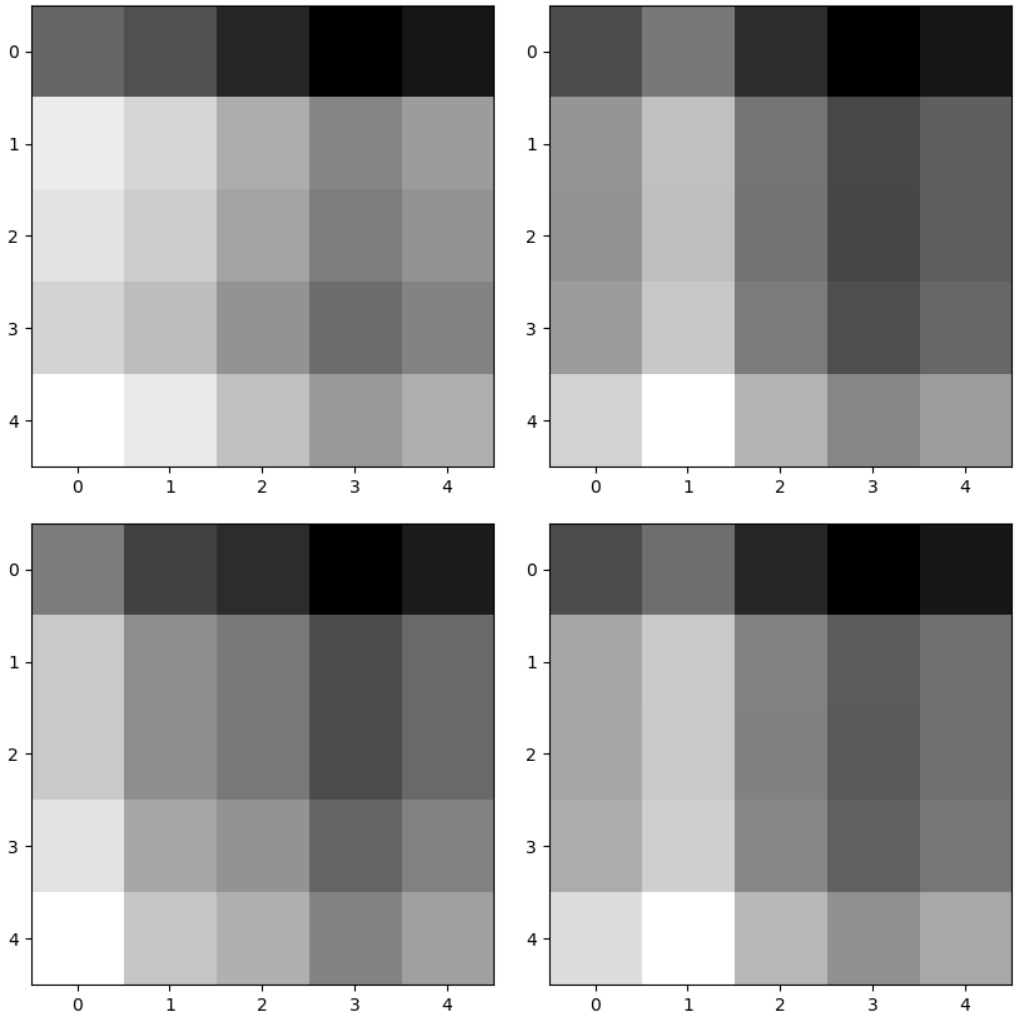
2. Plot of the mean cross-entropy loss as a function of the number of iterations.



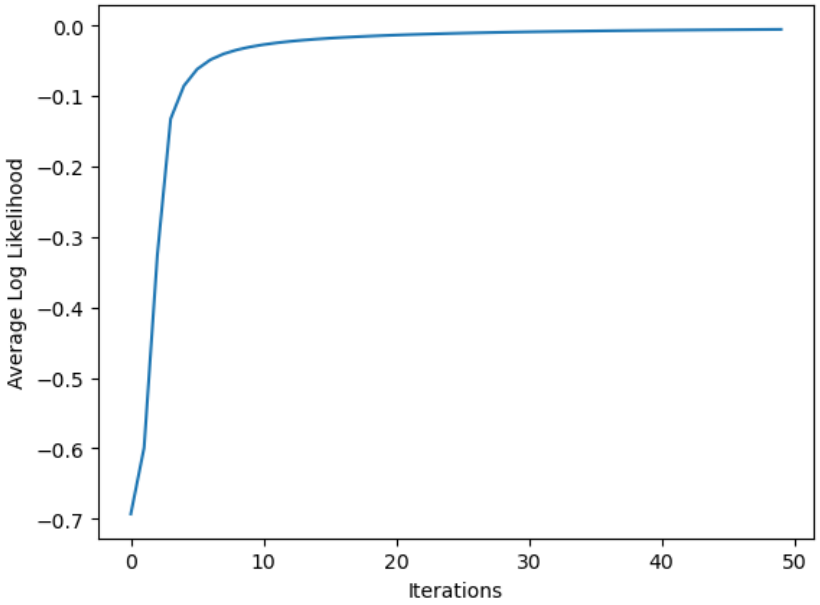
3. Hyper Parameters
   (a) Learning Rate = 0.01
   (b) Termination Criteria = 100 iterations
   (c) Regularization Term = 0.1
   (d) Kernel was initialized to values between -0.001 and 0.001.
   (e) Theta was initialized to values between -0.001 and 0.001.

5. CNN With Multiple Kernels

1. Image representations of the final kernels.



2. Plot of the Average Log Likelihood as a function of the number of iterations.



3. Hyper Parameters

   (a) Learning Rate = 0.1

   (b) Termination Criteria = 50 iterations

   (c) Regularization Term = 0.0001

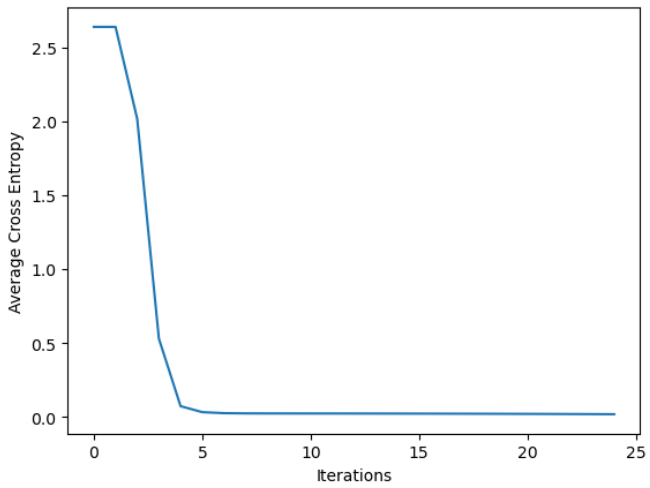
   (d) Kernel was initialized to values between -0.000001 and 0.000001.

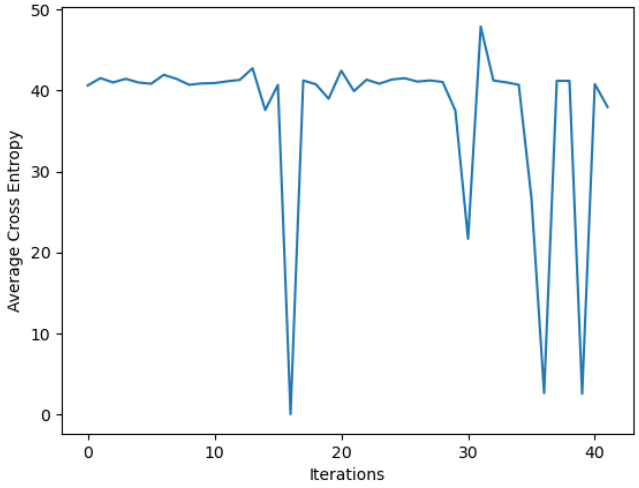
   (e) Theta was initialized to values between -1 and 1.

6. Multi-Kernel CNN For Image Classification

Note: For my first architecture in question 6, I have very low accuracy. I ran out of time, but could not figure out why the accuracy was poor. My kernel might have been poor. My training cross entropy graph looks fine however. For my second architecture, I tried my best, but could not successfully get a working architecture at all.

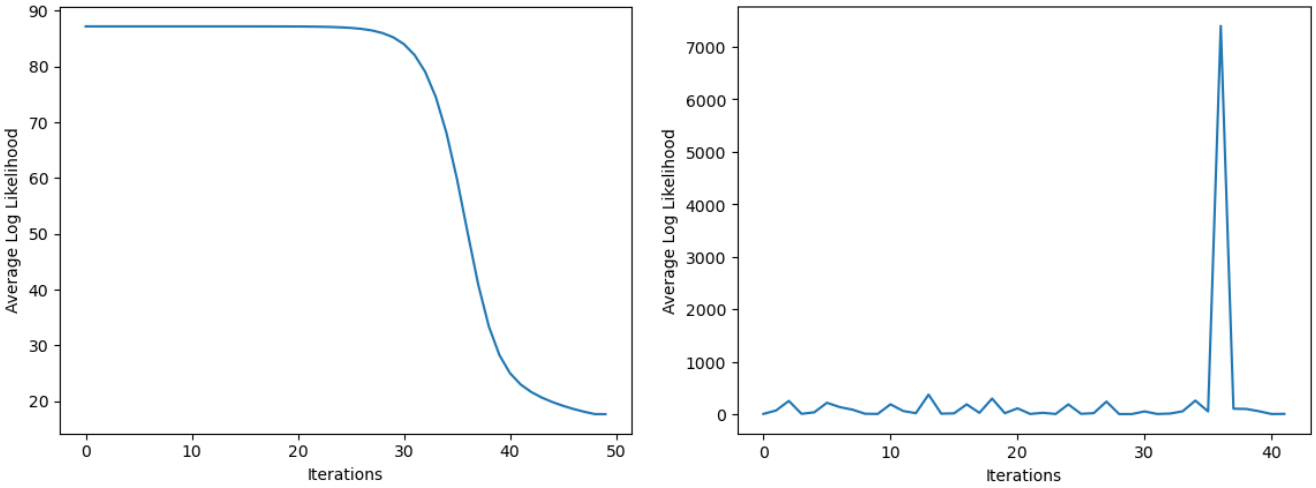
1. Architecture 1
   (a) A single 40 x 40 kernel
   (b) A max pooling layer of width 1 and stride 1
   (c) A flattened layer
   (d) Softmax and Cross Entropy Functions
2. Training and Testing Average Cross Entropy Graphs





3. Hyper Parameters
   (a) Learning Rate = 0.01
   (b) Termination Criteria = 25 iterations
   (c) Regularization Term = 0.1
   (d) Kernel was initialized to values between -0.000001 and 0.000001.
   (e) Theta was initialized to values between -0.001 and 0.001.
4. Training Accuracy = 100%, Testing Accuracy = 0%

1. Architecture 2
   (a) A single 40 x 40 kernel
   (b) A max pooling layer of width 1 and stride 1
   (c) A flattened layer
   (d) Linear and Squared Error Functions
2. Training and Testing Average Cross Entropy Graphs



3. Hyper Parameters
   (a) Learning Rate = 0.001
   (b) Termination Criteria = 50 iterations
   (c) Regularization Term = 0.0001
   (d) Kernel was initialized to values between -0.000001 and 0.000001.
   (e) Theta was initialized to values between -0.0001 and 0.0001.
4. Training Accuracy = 7.14%, Testing Accuracy = 7.14%