CS 615 Assignment3

Himanshu Gupta

May 2020

1 Theory

1.1

Given Matrix X = $\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix}$,Matrix K = $\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix}$

First we have to rotate K by 180 degrees to do X $*$ K, so our new K would be: $\begin{bmatrix}9 & 8 & 7\\6 & 5 & 4\\3 & 2 & 1\end{bmatrix}$

The matrix F would be of dimensions [size(X,1)-size(K,1)+1, size(X,2)-size(K,2)+1] = [3-3+1, 3-3+1] = [1,1]. To calculate F, we would be do element-wise multiplication between the X and this new rotated K, then sum all the elements. so, this would produce a single value F:

F = $[1*9+8*2+3*7+4*6+5*5+6*4+3*7+2*8+1*9] = [165]$

$$F_{ab} = (X_{ab} * K) = \sum_{i=-\frac{M}{2}}^{\frac{M}{2}} \sum_{j=-\frac{M}{2}}^{\frac{M}{2}} X_{(a-\mathrm{i},b-\mathrm{j})} K_{ij}$$

OR, F can calculated by using this formula:
where only possible a and b combination is a=2 and b=2, M = 3

1.2

Given Matrix F = $\begin{bmatrix}1 & 2 & 3 & 4 & 1 & 3\\4 & 5 & 6 & 0 & 0 & 12\\7 & 8 & 9 & 1 & 0 & 4\\-100 & -100 & -100 & -100 & -100 & -100\end{bmatrix}$

1.2.1 Max-Pooling width =2 stride = 2

In the max pooling, find the maximum values out of 2 X 2 matrix in F with moving by the stride of 2 in each iteration. The final matrix would be of dimensions [size(F,1)/width, size(F,2)/width] = [4/2, 6/2] = [2,3]

Matrix h = $\begin{bmatrix}max(1,2,4,5) & max(3,4,6,0) & max(1,3,0,12)\\max(7,-100,-100,8) & max(-100,-100,1,9) & max(-100,-100,0,4)\end{bmatrix} = \begin{bmatrix}5 & 6 & 12\\8 & 9 & 4\end{bmatrix}$

1.2.2 Mean-Pooling width =2 stride = 2

In the mean pooling, find the mean of the values in 2 X 2 matrix in F with moving by the stride of 2 in each iteration. The final matrix would be of dimensions [size(F,1)/width, size(F,2)/width] = [4/2, 6/2] = [2,3]

Matrix h = $\begin{bmatrix}12/4 & 13/4 & 16/4\\-185/4 & -190/4 & -196/4\end{bmatrix} = \begin{bmatrix}3 & 3.25 & 4\\-46.25 & -47.5 & -49\end{bmatrix}$

1.3

Find K in: X $*$ K = X

Now Let replace the convolution multiplication with the matrix product and replace the 'K' with 'Y'. Later to find the final solution, we just have to rotate our Y by 180 degrees.

In order to generalize our solution for all types of 'X' including non-inversible matrix (Not a square matrix), multiplying the both side by X^T as $X^T X$ would always be a square matrix.Then, the equation would be:

$X^T XY = X^T X$

$Y = (X^T X)^{-1}(X^T X)$

Using the matrix property of $A^{-1}A = I$ where I is an identity matrix, equation would be:

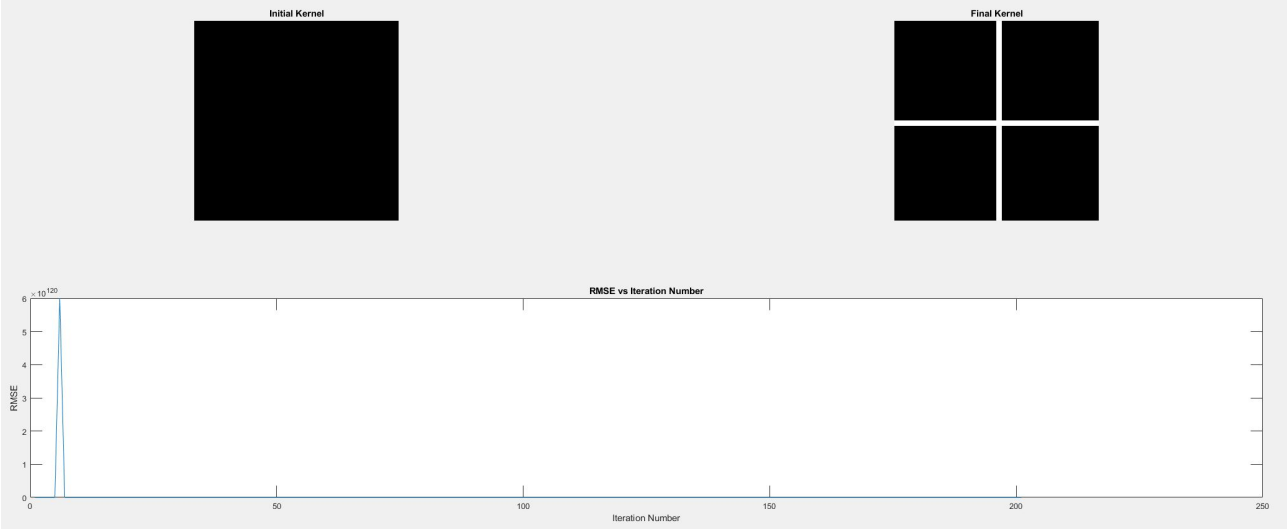
Y = I

Rotating the Identity matrix by 180 degrees would result in Identity matrix itself so,

K = I

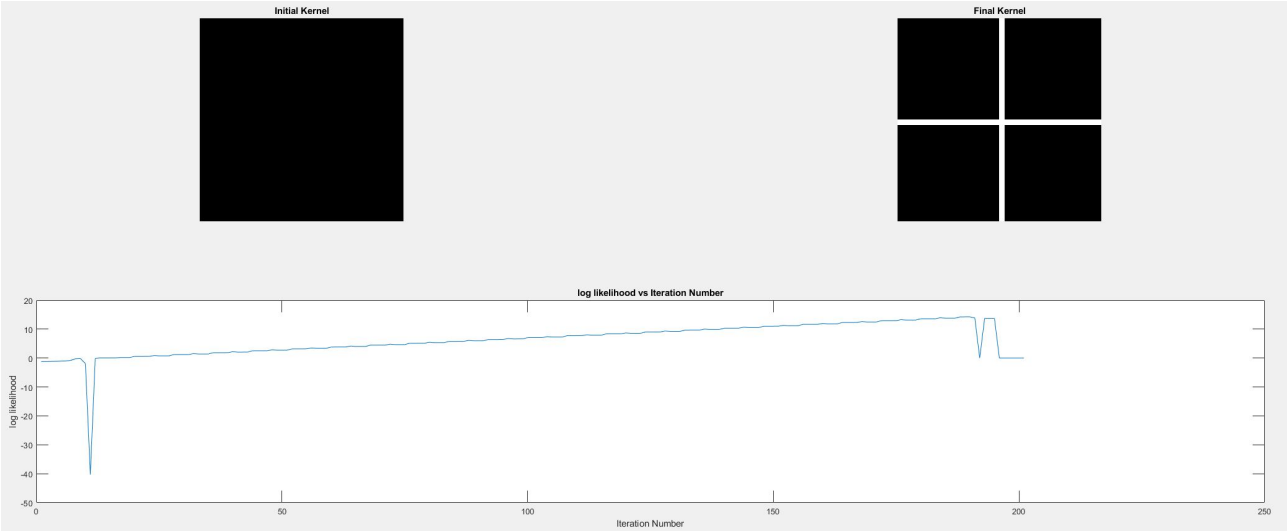
2 CNN for LSE Classification

Hyper-parameters:
L2 regularization = 0.000001
Learning rate = 0.55
Initialization of kernels and theta layer by randomly picking numbers from [0,0.000001]
Termination criteria is max iteration = 200



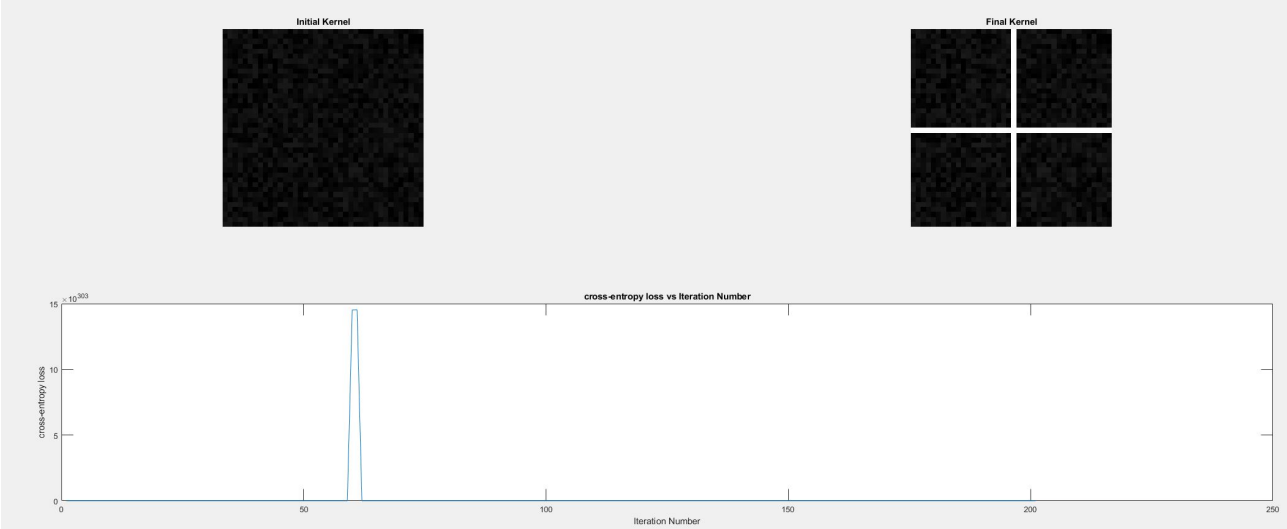
3 CNN for MLE Classification

Hyper-parameters:
L2 regularization = 0.001
Learning rate = 0.013
Initialization of kernels and theta layer by randomly picking numbers from [0,0.000001]
Termination criteria is max iteration = 200



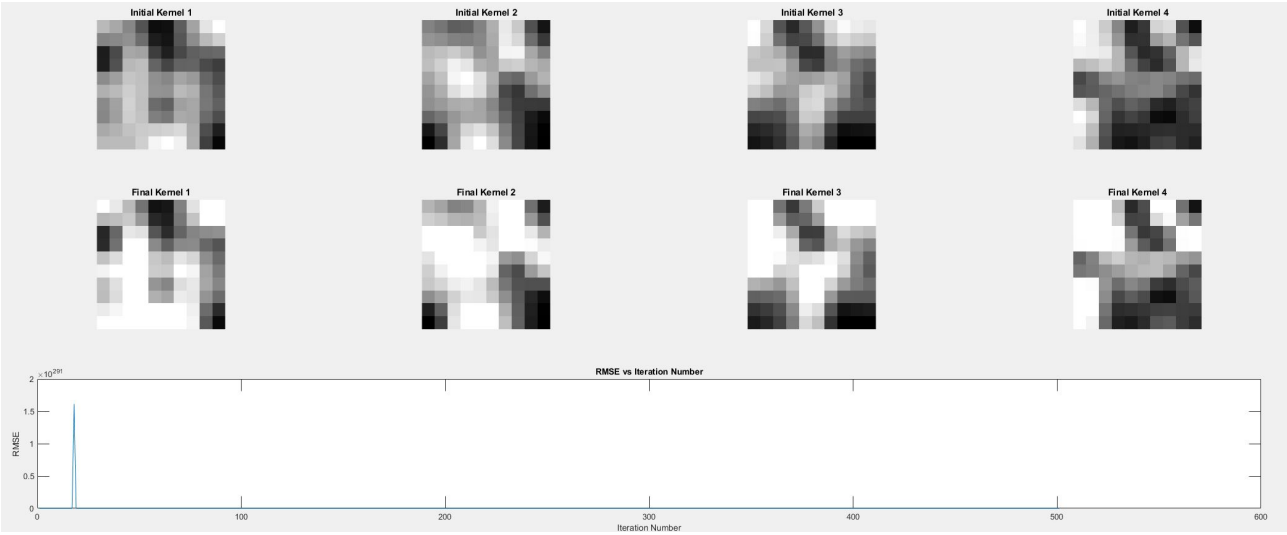
4 CNN for LCE Classification

Hyper-parameters:
L2 regularization = 0.001
Learning rate = 0.7
Initialization of kernels and theta layer by randomly picking numbers from [0,0.1]
Termination criteria is max iteration = 200



5 CNN With Multiple Kernels

Activation function is linear activation function
Objective function is RMSE (Root mean squared error)
Hyper-parameters:
L2 regularization = 0.1
Learning rate = 0.013
Initialization of kernels and theta layer by randomly picking numbers from [0,1]
Termination criteria is max iteration = 500

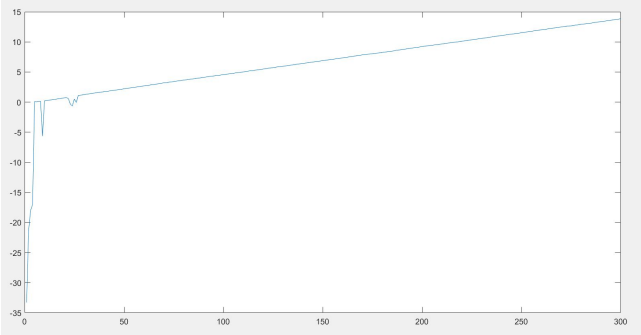


6 Multi-Kernel CNN For Image Classification

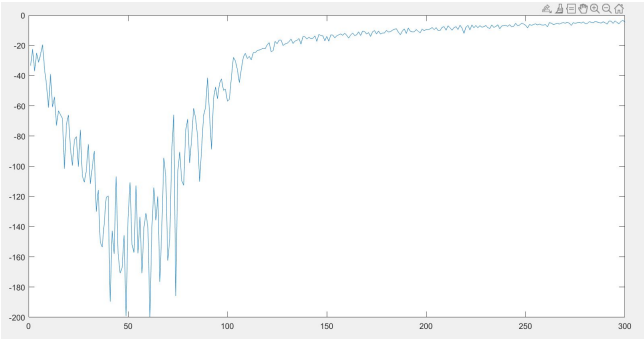
6.1 Activation function: logistic (sigmoid) function, Objective Function: log likelihood, max-pooling with width 2 and stride 2, and kernels dimension = 5 X 5 X 4

Hyper-parameters:
L2 regularization = 0.1
Learning rate = 0.02
Initialization of kernels and theta layer by randomly picking numbers from [0,1]
Termination criteria is max iteration =
Training Accuracy = 99.61%
Testing Accuracy = 79.61%

Log likelihood vs iterations for Training



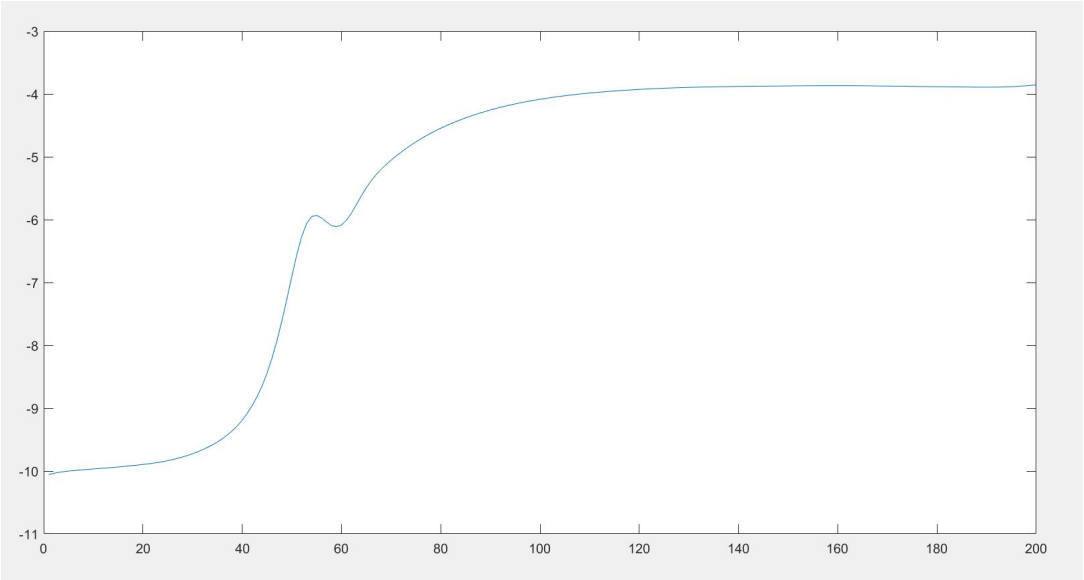
Log likelihood vs iterations for Testing



6.2 Activation function: logistic (sigmoid) function, Objective Function: log likelihood, max-pooling with width 2 and stride 2, and kernels dimension = 5 X 5 X 2

Hyper-parameters:
L2 regularization = 0.1
Learning rate = 0.01
Initialization of kernels and theta layer by randomly picking numbers from [0,0.01]
Termination criteria is max iteration =
Training Accuracy = 99.51%
Testing Accuracy = 80.58%

Log likelihood vs iterations for Training



Log likelihood vs iterations for Testing