CSCI 5525: Homework #3

Noah Hendrickson

Problem #1:

CSCI 5525 HW3

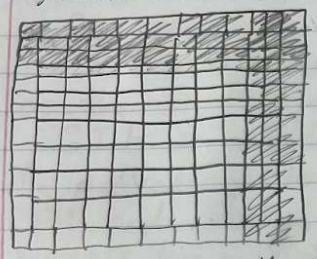
Noah Hendrickson

10x10 -> conv -> nxm -> flat -> kx1 -> full -> 10x1 -> out

Question #1:

Filter: 3x3 Input: 10x10 Stride: 3 Padding: 1

a) what are values of m, n, k



→ reduces down to a 4x4. Logic: Padding makes image a 12x12, Kernel is 3x3 with stride 3, so 4 kernels can fit width and 4 can fit height, thus, n = 4 and m = 4

As for k, flattening keeps same # of elements, just in a 1-D array, thus, $k = m \cdot n = 4 \cdot 4 = 16 = k$

b) Starting with W_{fc} and b, the input is 16x1, and output is 10x1.

W_{fc} input + b = output.

b's dimensions are trivial as they must be the same as output, (10x1)

W_{fc} has to turn 16x1 into 10x1. When doing matrix multiplication, columns of the first must match rows of the second. Rows of w will be the desired output rows. Thus, size of W_{fc} is (10x16) (10x16 · 16x1 -> 10x1)

As for W_{conv}, this is just (kernel size) x (channels) thus, (W_{conv} size is 3x3x1)

Problem #2:

Number of epochs that each was run for was 10. This was chosen because it seemed to be a good number for some and also I didn't want to take like an hour every time I wanted to see if it worked using the strategy that I used for the CNN and not being able to use GPU. In the case that the number of epochs is raised, the same process for convergence is taken as the CNN.

Optim: SGD	$\eta = 1e-5$	$\eta = 1e-4$	$\eta = 1e-3$	$\eta = 0.01$	$\eta = 0.1$
Train Loss	638.3525932	192.3993063	82.35496753	28.94842945	27.37736875
Train Error Rate	0.4878	0.1391	0.075	0.0277	0.0215

Optim: Adagrad	$\eta = 1e-5$	$\eta = 1e-4$	$\eta = 1e-3$	$\eta = 0.01$	$\eta = 0.1$
Train Loss	594.8198284	183.6154455	73.9978831	25.22639245	37.94906101
Train Error Rate	0.4288	0.1261	0.0666	0.026	0.0322

Optim: RMSprop	$\eta = 1e-5$	$\eta = 1e-4$	$\eta = 1e-3$	$\eta = 0.01$	$\eta = 0.1$
Train Loss	83.19246228	29.21636442	45.44201312	148.3480653	690.2229007
Train Error Rate	0.0752	0.0279	0.0274	0.0577	0.7867

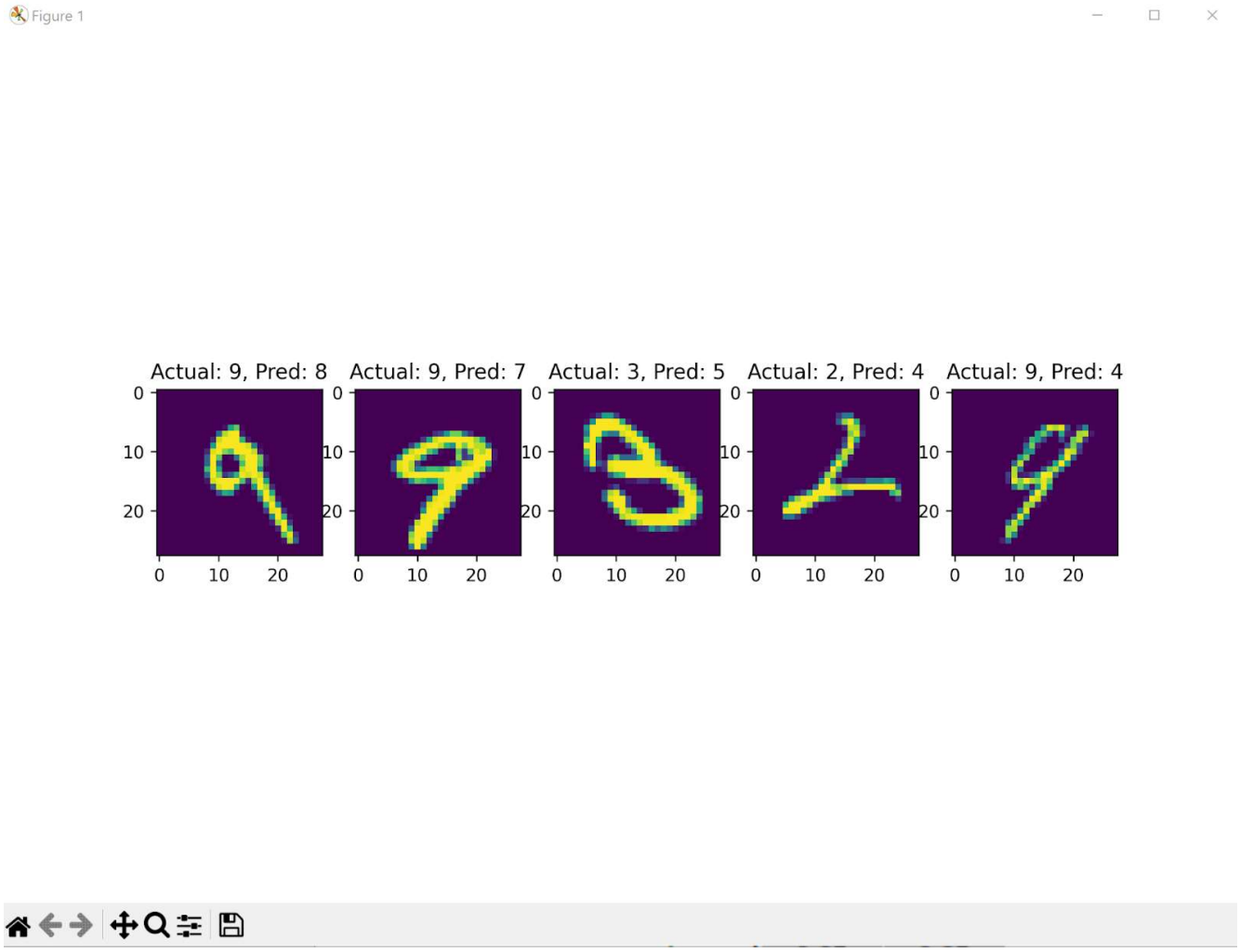
Optim: Adam	$\eta = 1e-5$	$\eta = 1e-4$	$\eta = 1e-3$	$\eta = 0.01$	$\eta = 0.1$
Train Loss	82.90945204	28.79559159	37.93969691	143.6300395	721.1442726
Train Error Rate	0.0754	0.0269	0.0249	0.0687	0.8972

Problem #3:

Number of epochs was variable for the CNN. The maximum number that the model could run for was input as 100. In the training stage, the train data was split into training and validation sets. If 5 epochs passed where the loss of the validation set was higher than the minimum seen loss on the validation set, the training ended early due to overfitting. Before returning epoch loss, the best model was restored.

Test Loss: 51.41966282718931
Test Error Rate: 0.0449

5 Incorrect:



The above images were likely misclassified because of their similarity to other numbers. The 9 on the far right could easily be mistaken for a 4 even by a person. The 2 is weirdly shaped and so is the 3. The two 9's on the left it's possible there were just not as many 9's in the dataset or they just look a little too unlike the other 9's such that they were misclassified.