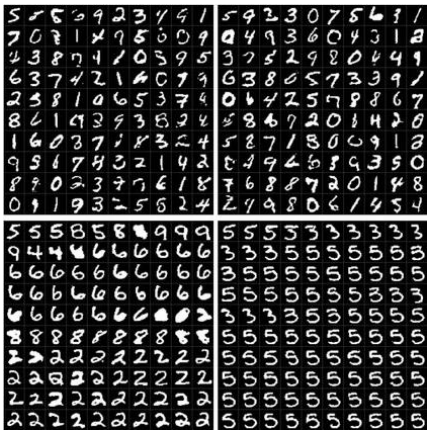
MSSE 277B: Machine Learning Algorithms

Homework assignment #9: Residual Neural Network
Assigned Apr. 6 and Due Apr. 18



1. Residual Neural Networks applied to classification. (20 pt) We will again use the MNIST data set to train, validation, and test but this time using a ResNN. As described in lecture, we are going to formulate a skip connection in order to improve gradient flow.

Using the CNN developed in HW#8, adapt your architecture to the one shown in the figure below (architecture with two layers each composed of one convolution and one pooling layer.) Use ReLU as your activation function. Use conv/pooling layers that with kernel, stride and padding size that lead to output size of 12x5x5 before flattening. Flatten the resulting feature maps and use two fully connected (FC) layers of output size (300,10). Add an additive skip connection from flattened layer to the second fully connected layer. Again, use the ADAM optimizer with learning rate of 1e-3, batchsize of 128, and 30 epochs (you can also train for longer if time permits). Split the MNIST training set into 2/3 for training and 1/3 for validation, you don't need to do KFold this time. Use batch normalization of data, choose some regularization techniques and converge your training to where the loss function is minimal.

(a) (10 pt) Run the model with and without batch normalization. Which give you better test accuracy?

(b) (10 pt) Run the model with and without the skip connection at learning rate of 5e-3 for 10 epochs. Do you see faster training and/or better test accuracy with the skip connection?