Exercise 3

Apply the CNN implementation with suitable node function from Exercise 1 to MNIST dataset, select one optimizer out of four (SGD, RMSProp, AdaGrad and Adam) to plot training loss and validation loss by epoch, training accuracy and validation accuracy by epoch, display the best result you have with suitable optimizer parameters and explain why you select that specific number to your optimizer?

Note:

1. MNIST database link: http://yann.lecun.com/exdb/mnist/, in MNIST dataset 60000 images for training and 10000 images for testing