Homework 3

Q1. Write the mathematical function for a neural network with 1 hidden layer and where the output layer is the (multidimensional) softmax operator. Using chain rule, find all the gradients with respect to the biases and weights.

Q2. Train the neural network from Q1 on the MNIST dataset using stochastic gradient descent or minibatch gradient descent. Report how you chose the number of neurons, if you used any regularization, and the accuracy.