CS 5316 – Natural Language Processing
Quiz 6 Solution
(Time limit: 12 minutes)

Design a feedforward neural network with 3 layers for document topic classification.
Suppose there are 5 topics and documents have varying lengths from 300 to 500 words.

   a. (2 points) Draw the architecture of the network showing the input layer, output
       layer, hidden layer(s), and connectivity between layers.
   b. (2 points) Describe the representation that you would use for documents.
   c. (1 point) Select the number of units in all the layers.
   d. (1 point) Fix the activation function in all the layers.
   e. (2 points) Define the parameters of the network. Be precise.
   f. (2 points) Describe briefly how you would train the networkincluding mentioning
       the type of loss function.

(a)
Input, layer 1, layer 2, output layer
Fully connected feedforward network from input to output layer
2 hidden layers

(b)
A number of representations can be used for documents.
   1. TF-IDF vector representation: TF-IDF calculated over the entire corpus; vector
       length is equal to vocabulary size, M
   2. Word embeddings: fix document length to first 400 words, concat word
       embeddings for these words in each document. If a document is shorter in length,
       pad with zeros.
   3. Document vectors: learn fixed length document embeddings, e.g., doc2vec or
       auto-encoder trained on TF-IDF vectors.

(c)
Input: M (doc vector length)
Hidden 1: 300
Hidden 2: 150
Output: 5

(d)
Hidden layers: tanh or ReLU
Output layer: soft max

(e)
Parameters include all weights and biases.
Input to hidden 1: 300 * M weights, 300 biases
Hidden 1 to hidden 2: 150 * 300 weights, 150 biases
Hidden 2 to output layer: 5 * 150 weights, 5 biases

(f)
Randomly initialize parameters
Feed each document through the network
Check the network's output with true label
If they differ update parameters such that categorical cross-entropy (loss function) is minimized
Repeat for all documents, and several epochs until convergence