CS 795: Deep Learning

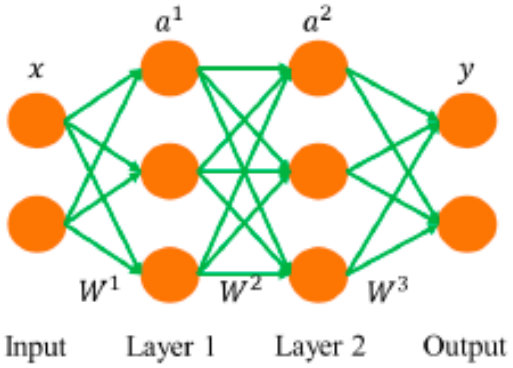
(Thanks to Dr. Reddy at Virginia Tech.)

Homework 1

Due Date: Feb 13th, 2019 (4:30PM) **Total: 100 Points**

Note: All implementations are required to be accomplished and submitted using the *Jupiter notebook (if using Python) or a demo file using other programming languages like Java/C*. Show your results and necessary comments in the notebook. *No handwritten homework is accepted.*

Consider the following deep (or shallow?) feedforward neural network with two hidden layers, two input dimensions and two outputs. You are provided two files in this assignment train.csv and test.csv



Assumptions:

a. We have zero biases in each layer, i.e., $z^1 = W^1 x$, $z^2 = W^2 a^1$, $z^3 = W^3 a^2$.
b. The activation function for two hidden layers are sigmoid function, i.e., $a^1 = \sigma(z^1)$ and $a^2 = \sigma(z^2)$, where $\sigma(z) = \frac{1}{1+e^{-z}}$.
c. The activation function for the output layer is a softmax function, i.e., $y_i = \frac{e^{z_i^3}}{\sum_j e^{z_j^3}}$.
d. The loss function is given by $L = \|y - \hat{y}\|^2 = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2$.

Questions:

1. Using backpropagation method and the computational graph learned in the class to derive $\frac{\partial L}{\partial W^1}, \frac{\partial L}{\partial W^2}, \frac{\partial L}{\partial W^3}$. Show the details of your work and solutions.
2. Implement the stochastic gradient decent (SGD) algorithm learned in the class to train the above neural network.
   a. The weights W^1, W^2, W^3 are randomly initiated.
   b. Define a learning rate η for the SGD and set it as a small number.
   c. In each training step, print the value of loss L. Print out the final weights.
   d. The training data has been provided in a separate file.

   In this question, you cannot call the optimization functions in the tensorflow, theano or other deep learning frameworks.

3. Train the above neural network on the same train data using any available deep learning framework. You are free to choose any package in any language. If you use python present your results as a python jupyter notebook.

4. Evaluate the models obtained in #2 and #3 with provided testing data (test.csv) using the metrics of precision, recall and f1-score. Provide a table comparing these approaches.

Deliverables:

- Go to courses.gmu.edu (blackboard) and submit a tar file with LastName_HW1.tar.gz
- The folder zipped should contain two folders:
- 1) PDF
  - Contains a typewritten pdf answering the questions above
- 2) src
  - Contains all source code written for the assignment.