

EE 580 Lab 3 Part4
Neural Nets
Summer 2017 Nazarian

Score:___/100 +
100 Extra
Credit

Student ID: _____

Name: _____

Assigned:

Due:

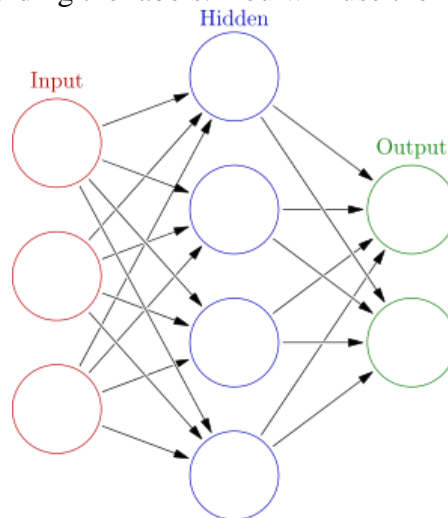
Late submissions will be accepted two days after the deadline with a maximum penalty of 15% per day: For each day, submissions between 12 and 1am: 2%, 1 and 2am: 4%, 2 and 3am: 8% and after 3am: 15%.

- **All assignments including this lab are based on individual work. No collaborations (including no discussions) are allowed.**
- **We may pick some students in random to demonstrate their design and simulations. Please watch the first lecture of this course regarding the academic integrity policies and also refer to the syllabus for a summary of AI policies (including the penalties for any violation).**
- **If you have any concerns or doubts about what is or is not allowed or prohibited in this course, please contact the instructor.**
- **ATTENTIONS: Start early otherwise you cannot finish this lab on time.**

PART A. Neural Nets in Machine Learning

Introduction

The goal of this assignment is to get some experience with text classification using the neural networks. You will be working with an email dataset and perform binary classification: SPAM or HAM (not spam). We are providing two sets of data. One is labeled, and one is specifically for testing and we are not providing the labels. You will use the labeled data to train your model.



An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in a brain. Here, each circular node represents an artificial neuron and an arrow

represents a connection from the output of one neuron to the input of another. This type of machine learning is very popular for classification.

More information:

Google

<http://karpathy.github.io/neuralnets/>

<http://cs231n.github.io/neural-networks-1/>

Requirements

Data Preparation: Based on the training files we provided, collect the number of occurrences in SPAM words and HAM words in each file. Specifically, you can define SPAM words like fraud, pill and so on and calculate occurrences of SPAM words. To calculate HAM words, just subtract the number of occurrences in SPAM words from the total number of words in each file. In this way, each file is characterized as a 2-tuple (i, j) where i and j is the number of occurrences in SPAM and HAM words. Then you can use this dataset as actual training data to feed into your neural net.

Neural Networks: Look at simple.py posted on BB and this website for installation and more explanations. <https://nimblenet.readthedocs.io/en/latest/index.html>. You need to train and validate your network using dataset you generated. It is your choice to choose the number of hidden layers and the number of neurons in each hidden layer.

Testing: Based on testing files, collect the number of SPAM words and HAM words in each file and use the neural net you just trained to predict whether a testing file is a SPAM email or a HAM email.

PART B. Impact of Hidden Layers

Find a suitable data set that would show the impact of number of hidden layers and the number of nodes in each layer. This part has a maximum of 100 extra credit points, depending on the quality of your work.

Submission

1. Zip all the files you need to submit into a zip file named: "firstname_lastname_lab3_part4".zip.
2. Files given, 32 bit IEEE format adder, multiplier and divider.
3. Your zip file should include Readme.pdf, Verilog files for **module SOP, module Sigmoid, module adder, module multiplier, module divider**. You can create a single file named

Activation.v under **module Activation**. Make sure it works.

4. In your Readme.pdf,
 - a. **Create the image of the architecture you designed for the pipelined model in part A.**
 - b. **Specify if you used a different sigmoid in part b. Please mention the formula for it.**
 - c. Include any information that you think the course staff, especially the grader should know while grading your assignment: references, any non-working part, any concerns, etc.
 - d. The citations should be done carefully and clearly, e.g.: *“to write my code, lines 27 to 65, I used the Dijkstra's shortest path algorithm c++ code from the following website: www.SampleWebsite.com/...”*
 - e. The Readme file content of labs and PAs can be hand-written or typed. In case you decide to hand-write, then please scan and include in your Readme.pdf.
- NOTE: this policy is different from that of the HW. For HW assignments, the solutions have to be handwritten.

Use the provided BB submission link to submit your zip file for this assignment