EE 526X Deep Machine Learning: Theory and Practice — Homework 1
Assigned: 09/11. Due: 09/25

Problem 1. The Spambase Data Set contains email spam data for 4601 email messages. Download the data from `https://archive.ics.uci.edu/ml/datasets/spambase` and divide the data into training set and test set. The training set should contain the first 2/3 of spam and ham (i.e., non-spam) messages. The test set should contain the last 1/3 spam and ham messages.

(a) Write a logistic regression program (function) using gradient descent algorithm. And train the weights using training set and then test the result on the test set. Experiment with the step size (learning rate).

(b) Next, normalize the features, so that each feature in the training data has mean 0 and variance 1. Then run logistic regression on the normalized data.

You need to submit the source program, in one the three languages: Python (preferred), MATLAB/Octive, and C. Also the results on the learning rate used, and training and test errors should be reported.

Problem 2. Write a program to simulate a three layer neural network that has the following specifications:

(a) Input dimensions: 57 (all real numbers);

(b) Output dimension: 1 (binary);

(c) Layer 1: 5 neurons, with ReLU (rectified Linear Unit) nonlinearity;

(d) Layer 2: 3 neurons, with ReLU nonlinearity;

(e) Output Layer: 1 neuron, with logistic nonlinearity;

(f) The objective function is cross-entropy (logistic regression).

Use the same training and test data from the Spambase Data Set. Train the neural network using forward and backward propagation and gradient descent. You need to actually implement forward and backward propagation in the program.

You need to submit the source program, in one the three languages: Python (preferred), MATLAB/Octive, and C. Also the results on the learning rate(s) used, and training and test errors, and running time should be reported.

```
END OF ASSIGNMENT
```