EE 526X Deep Machine Learning: Theory and Practice — Homework 01
Assigned: 09/07. Due: 09/17

Problem 1. The `cars` dataset contains a set of 50 data points (see file `cars.csv`). Each data point is a triplet, where the first component is the index, the second component is the speed, and the third component is distance. The observations give the speed of cars and the distances taken to stop. Write a `Python` program that fit an n-th order polynomial to the data. That is, model the distance as a n-th polynomial of speed. Find the coefficients for the cases $n = 1, 2, 3, 4$. And plot points as well as the fitted model in one plot (including all data points and the 4 polynomials. Note that you should write the regression program yourself, instead of using any existing packages.

Problem 2. Use `Python` to implement the perceptron algorithm and test it on the following data: where there are 6 points, (x_1, x_2) is the input, and y is the binary output

	x_1	x_2	y
1	1	2	1
2	1	4	1
3	2	2	1
4	4	2	-1
5	3	4	-1
6	2	3	-1

label. Initialize your algorithm with the vector

$$\theta = [b, w_1, w_2]^T = [0, 0, 0]^T. \tag{1}$$

Plot the points and the final hyperplane (a line) on the same graph.

Problem 3. 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 `Python` (preferred) or `MATLAB/Octive`. Also the results on the learning rate used, and training and test errors should be reported.

Problem 4. Prove that the softplus function

$$f(z) = \log(1 + e^z) \tag{2}$$

is convex in z by showing that its second derivative is positive for all z.

END OF ASSIGNMENT