ECGR 3090/4090 - Introduction to Machine Learning

Due: 11:59PM, 10/05, 2020 (Monday, mid-night)

Submission Guidelines:

1. Prepare a report showing your (detailed) solution to each problem. Scan (PDF) your solution if you write on a piece of paper.

2. For programming problems, submit your code as well. If there are multiple programming problems, please properly name your code files corresponding to each problem (e.g., P1_gradescent.m). You can also organize your codes in separate folders for each problem and submit a "zip" file.

Homework #1 Linear regression (2 problems)

Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question. For programming problems, you can use any programming language, e.g. MATLAB, python, etc.

[25 points] **Problem 1:**

We will use a dataset provided: "D3.csv"

Let the first three columns of the data set be separate explanatory variables x_1, x_2, x_3. Again, let the fourth column be the dependent variable y.

Run linear regression simultaneously using all three explanatory variables. Report the linear model you found by running the **gradient descent** algorithm. Predict the value of y for new (x_1, x_2, x_3) values (1, 1, 1), for (2, 0, 4), and for (3, 2, 1). (Note: You **cannot** use built-in function from ML libraries for gradient descent, you have to implement it yourself.)

Attach your implementation code as well! For example, assume I code two functions to solve this problem (this is only an example, you don't have to follow this strictly):

Function1.m

```
function run_linear_regression ()
    % load data
    xxxx
    % gradient decent
    xxxxx
end
```

Function2.m

```
function gradient_decent ()
% this is an implementation of gradient decent
   xxxxxxx
end
```

[25 points] **Problem 2:**

A website specializing in dongles (dongles-r-us.com) wants to predict the total dollar amount that visitors will spend on their site. It has installed some software that can track three variables:

• time (the amount of time on the page in seconds): x_1,

• jiggle (the amount of mouse movement in cm): x_2, and

• scroll (how far they scroll the page down in cm): x_3.

Also, for a set of past customers they have recorded the

• sales (how much they spend on dongles in cents): y. We see a portion of their data set here with n = 11 customers:

time: x_1	jiggle: x_2	scroll: x_3	sales: y
232	33	402	2201
10	22	160	0
6437	343	231	7650
512	101	17	5599
441	212	55	8900
453	53	99	1742
2	2	10	0
332	79	154	1215
182	20	89	699
123	223	12	2101
424	32	15	8789

Let the first three columns of the data set be separate explanatory variables x_1, x_2, x_3, and the fourth column be the dependent variable y. Compute the closed-form solution (or analytical solution) for the hypothesis function parameters: $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2, \theta_3]$. Show each step. (You can use tools and built-in functions e.g. MATLAB to compute the matrix inverse in deriving your solution.) You can submit your code as well.