**[CS-4267 / 5267 Spring 2024] Deep Learning**

**Assignment 1: Linear Regression**

## I. Purpose:

This homework contains classic regression tasks for linear and logistic regression, which is a warm-up project for this class. **Please finish this home using Google Colab**.

* Learn how to use CoLab.
* Learn how to load customized data.
* Learn how to perform gradient descent for linear regression.
* Learn how to plot the intermediate results.

II. Download Data

food\_truck\_data.txt: <https://drive.google.com/file/d/1YcPg7-o3sqZHT5Hzyj0ekDs_PsRqQ1fK/view?usp=sharing>

housing\_price\_data.txt:

<https://drive.google.com/file/d/17VAHegukchKyIOdoPkil8oDyMzbmg1W0/view?usp=sharing>

## II. Description

Task 1: Linear Regression with One Variable

Suppose you are the CEO of a restaurant franchise and are considering different cities for opening a new outlet. The chain already has food trucks in various cities and you have data for profits and populations from the cities. You'd like to figure out what the expected profit of a new food truck might be given only the population of the city that it would be placed in. Predict the profits for a food track using simple regression based on the available data.

The file food\_truck\_data.txt contains the dataset for the problem. The data consists of two columns; the first column is the population of a city and the second column is the profit of a food truck in that city. A negative value for profit indicates a loss.

* Fit a linear model to the data (detailed grading terms please see Grading and Submission)
  + Compute the parameters of the model using gradient descent
  + Compute and plot the cost function as you perform the gradient descent iterations
* Plot the data with the linear model (detailed grading terms please see Grading and Submission)

Task 2: Linear Regression with Multiple Variables

Suppose you are selling your house and you want to know what a good market price would be. You can use available data on recent housing prices sold and learn a linear model.

The file housing\_price\_data.txt contains the dataset for the problem. The data consists of three columns; the first column is the size of the house (in square feet), the second column is the number of bedrooms, and the third column is the price of the house.

* Fit a linear model to the data (detailed grading terms please see Grading and Submission)
* With the following requirements (detailed grading terms please see Grading and Submission)
  + Use feature scaling
  + Compute the parameters using gradient descent
  + Compute and plot the cost function as you perform the gradient descent iterations

## III. Grading and Submission

* The assignment will be evaluated in a total of 50 points. The basic scores are generally given based on the following table.

| Using Google Colab (5 points)   * Submit a working Colab ipynb file (2.5’) * With required results embedded in the ipynb 2.(5’)   Part 1: Linear regression with one variable (20 points)   * Define a linear model to the data (3’) * Explain the model you defined with >=50 words (e.g., which parameters are the inputs, what are the output, what are the ones to compute) (3’) * Compute the parameters of the model using explicit gradient descent (4’)   \* Please don’t use a closed-form solution or a non-interactive solution.   * The gradient descent is formed as iterative loops (4’) * Plot the cost function as you perform the gradient descent iterations (3’) * Plot the data with the linear model (3’)   Part 2: Linear regression with multiple variables (25 points)   * Use feature scaling (5’) * Define a linear model to the data (5’) * Explain the model you defined with >=50 words (e.g., which parameters are the inputs, what are the output, what are the ones to compute) * Compute the parameters of the model using explicit gradient descent (5’)   \* Please don’t use a closed-form solution or a non-interactive solution.   * The gradient descent is formed as iterative loops (5’) * Plot the cost function as you perform the gradient descent iterations (5’) |
| --- |

For each 5’ scale.

5’ = perfectly correct

4’ = minor flaw

3’ = mostly incorrect

2’ = totally incorrect

1’ = do something

0’=not do anything

* The assignment should be submitted with two files:
  + 1. A single PDF report file should be submitted to Brightspace with last name and VUID (e.g., “Huo\_huoy1.pdf”). The ideal PDF report file is a printed Colab ipynb file with required results embedded.
    2. All source code should be submitted to Brightspace as a single zip file with last name and VUID (e.g., “Huo\_huoy1.zip”).
* The deadline of submission is on the course website <https://hrlblab.github.io/CS4267.html>

Appendix (some notes)

1. Example code for this homework

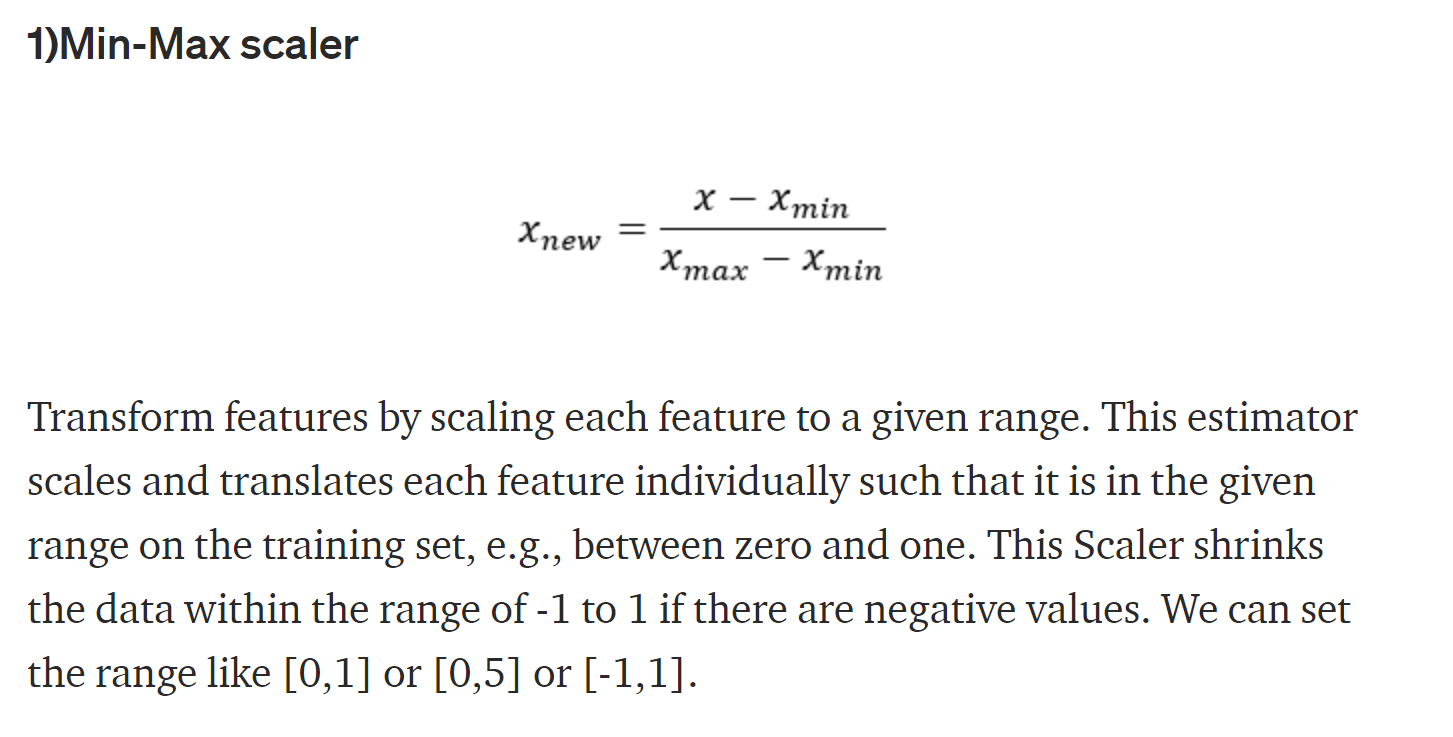
<https://www.cs.toronto.edu/~frossard/post/linear_regression/>

You can refer to this implementation. (But please do not copy-paste the code)

1. For feature scaling, the range of inputs in the multi-variable changes a lot. So we use feature scaling to normalize the inputs. There are many good options of feature scaling

<https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35>

The simplest is the min-max scaler



1. The cost curve is the cost-iteration figure, like

