CMPE 258, Spring 2018

Assignment #1

Due 11:59pm on Thursday, February 1st, 2018

Notes

This programming assignment should be submitted in Canvas as a format of ipython notebook (assignment_1_yourFirstName_LastnName.ipynb).

You can discuss how to solve the problem with other students or search internet or other resources, but the work should be your own.

The submitted ipynb should be executable without any extra work.

1 (10pts). Linear regression with one variable from scratch

Using Jupyter notebook, load the data (ex1data1.csv). Visualize data using scatter plot.
The first column is Population of City in 10,000s, and the second column is profit of food truck in 10,000.
In order to predict the profit, fit the data using gradient descent method (without matrix). You need to calculate cost function and update weight using gradient descent method. Try several different learning rate. Please print Root Mean Squared Error (RMSE) after optimization.

2(30pts). Linear regression with multiple variables from scratch

Using Jupyter notebook, load the data (ex1data2.csv). Visualize data.
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.
In order to predict the housing price, fit the data using gradient descent method (without matrix). You need to calculate cost function and update weight using gradient descent method. Try several different learning rate. Please print the Root Mean Squared Error (RMSE) after optimization.

2-1. Linear regression with multiple variables using matrix

Fit the data (ex1data2.csv) using matrix calculation. You need to calculate cost function and update weight. Please print the Root Mean Squared Error (RMSE) after optimization.

2-2. Linear regression with multiple variables using Normal equation

Fit the data (ex1data2.csv) using Normal equation. You need to calculate cost function and update weight. Please print the best Root Mean Squared Error (RMSE) after optimization.

3(60pts). Linear regression with multiple variables

Using Jupyter notebook, load the data (ex1data3.csv).
This is California housing dataset. The original database is available from http://lib.stat.cmu.edu
The data contains 20,640 observations on 9 variables. This dataset contains the average house value as target variable and the following input variables (features): average income, housing average age, average rooms, average bedrooms, population, average occupation, latitude, and longitude
(R. Kelley and Ronald Barry, Sparse Spatial Autoregressions,\nStatistics and Probability Letters, 33 (1997) 291-297) .

3-1. Linear regression with multiple variables using matrix

Fit the data (ex1data3.csv) using matrix calculation. You need to calculate cost function and update weight.
Please print the Root Mean Squared Error (RMSE) after optimization.
.

3-2. Linear regression with multiple variables using Normal equation

Fit the data (ex1data3.csv) using Normal equation. You need to calculate cost function and update weight.
Please print the best Root Mean Squared Error (RMSE) after optimization.

3-3. Linear regression with multiple variables using scikit-learn linear regression model

Fit the data (ex1data3.csv) using linear regression from scikit-learn library. You need to calculate cost function and update weight.
Please print the best Root Mean Squared Error (RMSE) after optimization.

3-4. Linear regression with multiple variables using TensorFlow

Fit the data (ex1data3.csv)using linear regression using TensorFlow. Please do not use Normal equation TensorFlow. You need to calculate cost function and update weight.
Please print the best Root Mean Squared Error (RMSE) after optimization.