CS 688

Fall 2016

Homework 2.

Due 10/4

1. Linear regression with one variable: In this part of this exercise, you will implement linear regression with one variable to predict profits for a food truck. Suppose you are the CEO of a restaurant franchise and are considering different cities for opening a new outlet. The chain already has trucks in various cities and you have data for profits and populations from the cities. You would like to use this data to help you select which city to expand to next. The file ex1data1.txt contains the dataset for our linear regression problem. 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.
    a. Before starting on any task, it is often useful to understand the data by visualizing it. For this dataset, you can use a scatter plot to visualize the data, since it has only two properties to plot (profit and population). Write a script that does that.
    b. Write a script to fit the parameters of a linear model to predict profits from population. Have your script show the change in the loss function vs. the number of iterations.
    c. Write a script that plots the function $L(w)$ vs. the two parameters.
    d. Repeat b. with 2/3 of the data. Use the other 1/3 to test your model. Evaluate $L(w)$ in the training and test data sets.
2. Linear regression with multiple variables: In this part, you will implement linear regression with multiple variables to predict house prices. The file ex1data2.txt contains a training set of housing prices in Portland, Oregon. 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.
    a. Write a script that learns the parameters of the model using Gradient Descent.
    b. Write a script that solves the problem using the Normal Equation.
    c. Compare the results.
    d. Repeat a. and b. with 2/3 of the data. Use the other 1/3 to test your model. Evaluate $L(w)$ in the training and test data sets.
3. Problem 2.9 Rogers book
4. Problem 3.2 Rogers book