Homework 1

Fall 2022 CMSC 478-01 Introduction to Machine Learning

Instructor: Abhishek Kumar Umrawal

Maximum Points: 100
Due Date: Oct 13, 2022, 11:59 PM

Instructions

1. Answer **ALL** questions.
2. Show all the work.
3. Submit your work as a Jupyter notebook.
4. Report the results in a presentable manner. Just computer outputs will not carry any points.

Questions

1. Consider the dataset `mileage.xlsx` which has observations for 32 different models of cars on mileage and 10 other features affecting the mileage. Perform the following tasks. [50+20+5+20+5]

   (a) Learn a multiple linear hypothesis to predict the car mileage based on the available features using gradient descent algorithm. Take learning rate as 0.02.

   (b) Repeat (a) for the following choices of learning rate: 0.001, 0.003, 0.01, 0.03, 0.1, 0.3 and 1. For every choice of the learning rate, make a plot that shows how the cost function changes over the successive iterations of the gradient descent algorithm.

   (c) What can you say about the relationship between learning rate and rate of learning based on your solution to (b)?

   (d) Learn a multiple linear hypothesis to predict the car mileage based on the available features using normal equations method.

   (e) Compare the hypotheses learnt in (a) and (d). Provide explanation of your observation based on this comparison.

   Data Source: Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.

   Note: For all your implementations of the gradient descent algorithm, define *convergence* if the absolute difference between the new and the old iterate is less than 10^{-4}.