Written by: Professor Wu, Solutions by:
DS 4420: Machine Learning II

Assignment 2

Question 1. Given the following data (10pts)

$$\begin{bmatrix} x & y \\ 1 & 1 \\ 2 & 1 \\ 1.5 & 0 \\ 3 & 2 \end{bmatrix}.$$

Assume that the function is a parabola where

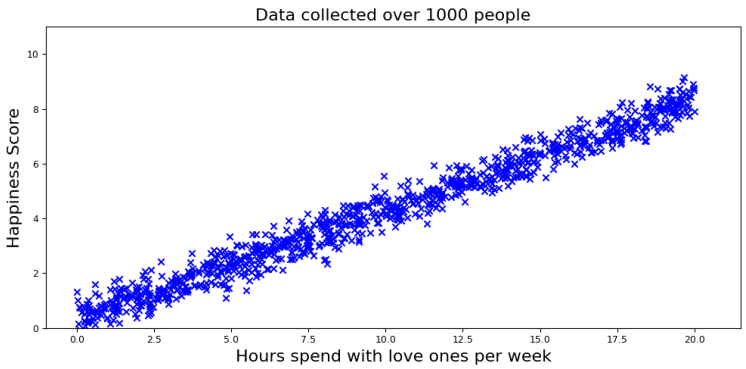
$$f(x) = ax^2 + bx + c.$$

1) Write the regression code with gradient descent to identify a function $f(x)$ such that given an x value, it returns the best prediction of y.
2) Once you have your own function, plot the 4 points and along with the line as predicted by your function $f(x)$.

Question 2. (30pts) (Use Python for this question) You went around the campus and asked 1000 people 2 questions.

- How many hours a week do you spend with the people you truly love?
- What is your happiness level out of 10?

After you collected the data, you plotted it out and it looks like the included figure.



This data can be found in

- **time_with_loved_ones.csv**
- **happiness.csv**

1) Use linear regression by assuming that the data can be described by the function

$$f(x) = af_1(x) + bf_2(x).$$

where

$$f_1(x) = x \quad \text{and} \quad f_2(x) = 1.$$

Write your own gradient descent algorithm with Python code to find the best a and b values.
2) Plot out all the points in blue along with the equation $f(x)$ in red.
3) According to your equation $f(x)$, if you spend 15 hours a week with your loved ones, how happy would you be?

Question 3. (20 points) Load the two csv files

`easier_data.csv`
`label.csv`.

This data predicts tomorrow's stock price difference given the previous day's data.

1) Use Preprocessing on the data
2) Perform regression on this dataset
   (a) Solve it with Sklearn linear regression
   (b) Solve it by writing your own gradient descent
   (c) Solve it with the closed-form solution
   (d) Print out the final total error
   (e) Print out the prediction of your function for every sample next to the true label
   (f) Compare them against the true labels. How good is your prediction?
3) Repeat the previous parts using Polynomial regression (2nd order)

Question 4. (20 points)
Given the data

$$\begin{bmatrix} x & y \\ 0 & 1 \\ 1 & 0 \\ 2 & 2 \\ 3 & -2 \end{bmatrix}.$$

If we assume that the function to predict y from x is a linear function then the function would like.

$$f(x) = a(x) + b \tag{1}$$

The goal is to use the data to identify the best a and b using the Closed-Form solution. Solve the question by hand as well as using Python.

Question 5. (20 points)

Let's say we have some $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and the feature maps are

1) $\varphi_1(x) = x_1$
2) $\varphi_2(x) = x_2$
3) $\varphi_3(x) = x_1 x_2^2$
4) $\varphi_4(x) = x_1^3$
5) $\varphi_5(x) = x_2^2$
6) $\varphi_6(x) = 1$

Given the following data

$$\begin{bmatrix} x_1 & x_2 \\ 0 & 2 \\ 2 & -1 \\ -2 & 1 \\ 3 & 3 \\ 4 & 1 \end{bmatrix}.$$

Transform the data using the feature map