Linear Regression (Univariate)

1. Consider the problem of predicting sunny weather condition in each week of 2023 given the sunny weather condition in each week of 2022.

   In this scenario, x represents the number of days in each week that the weather is sunny in 2022. The value of y is defined as “the number of sunny days” in each week of 2023 which we want to predict. The following training set is a sample of few weeks with number of sunny days in each of them. We use *m* to denote the number of training examples.

   Write the hypothesis equation $h_\theta(x\ )$ for linear regression on the dataset below.

x	y
3	2
2	3
4	5
1	1
5	4

2. For the training set given above, what is the value of m?

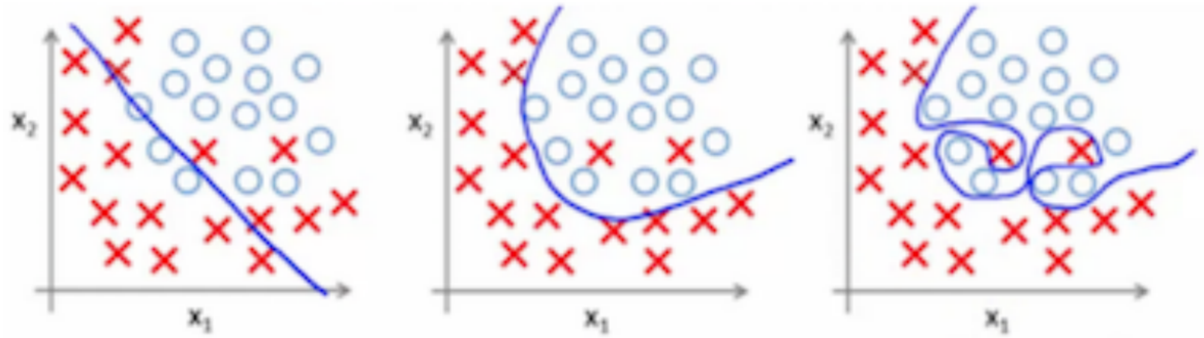
3. For this question, continue using the data provided in (1). Recall the definition of cost function for linear regression is

$$J(\theta_0, \theta_1) = \frac{1}{2m} * \sum_{i}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

   What is *J*(0,1)?

4. Suppose we set $\theta_0 =- 1,\ \theta_1 = 0.5,$ what is $h_\theta(3)$? (Bonus question: Calculate mean absolute error for the given parameter in question 4)

5. Three different classifiers are trained on the same data. Their decision boundaries are shown below. Which of the following statements are true?



The leftmost classifier has high robustness, poor fit.
The leftmost classifier has poor robustness, high fit.
The rightmost classifier has poor robustness, high fit.
The rightmost classifier has high robustness, poor fit

Linear Regression (Multivariate)

1. Suppose we have m = 4 houses, and each house has an area and number of bedrooms which can be used to predict the house price. A dataset of the features is as follows:

	Bedrooms	Sqft_area	Price
1.	1	880	490,000
2.	3	1930	630,000
3.	4	1940	640,000
4.	3	1350	570,000

You'd like to use polynomial regression to predict a house price from its numbers of bedrooms and sqft_area. Concretely, suppose you wish to fit a model of the form.

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

where x_1 is the number of bedrooms and x_2 is sqft_area. Further you plan to use both feature scaling (dividing by the max-min, or range, of a feature) and mean normalization.

What is the normalized feature x_2^4(i.e. for the fourth training data)?

2. You run gradient descent for 12 iterations with α = 0.2 and compute J(θ) after each iteration you find that the value of J(θ) increases over time. What would you do to correct this issue?

3. Suppose you have m = 23 training examples with n = 5 features (excluding the additional all-ones feature for the intercept term, which you should add). The normal equation is $\theta=(X^TX)^{-1}X^T y$. For the given values of m and n, what are the dimensions of θ, X, and y in this equation?

    1. X is 23 × 6, y is 23 × 6, θ is 6 × 6

    2. X is 23 × 5, y is 23 × 1, θ is 5 × 1

    3. X is 23 × 6, y is 23 × 1, θ is 6 × 1

    4. X is 23 × 6, y is 23 × 1, θ is 5 × 5

4. Suppose you have a dataset with m = 1000000 examples and n = 200000 features for each example. You want to use multivariate linear regression to fit the parameters θ to our data. Should you prefer gradient descent or the normal equation?

5. Which of the following statements are true?
    1. MAE doesn't add any additional weight to the distance between points. The error growth is linear.
    2. MSE errors grow exponentially with larger values of distance. It's a metric that adds a massive penalty to points that are far away and a minimal penalty for points that are close to the expected result.
    3. It is necessary to prevent the normal equation from getting stuck in local optima.
    4. It prevents the matrix XTX (used in the normal equation) from being non-invertable (singular/degenerate).