Linear Regression

Simple Linear Regression, Multiple Linear Regression, Least Square Gradient Method

by Dr. Tumpa Banerjee

Introduction

- The objective of regression is to determine the value of one or more continuous target variables Y based on the d-dimensional input vector X. Given a set of N observations $\{X_n\}$, together with the target values $\{Y_n\}$.
- The objective is to predict the value of y for an input x.
- A linear regression model assumes that the regression function $E(Y|X)$ is linear in the inputs $X_1, X_2, \ldots, X_p$.

Introduction

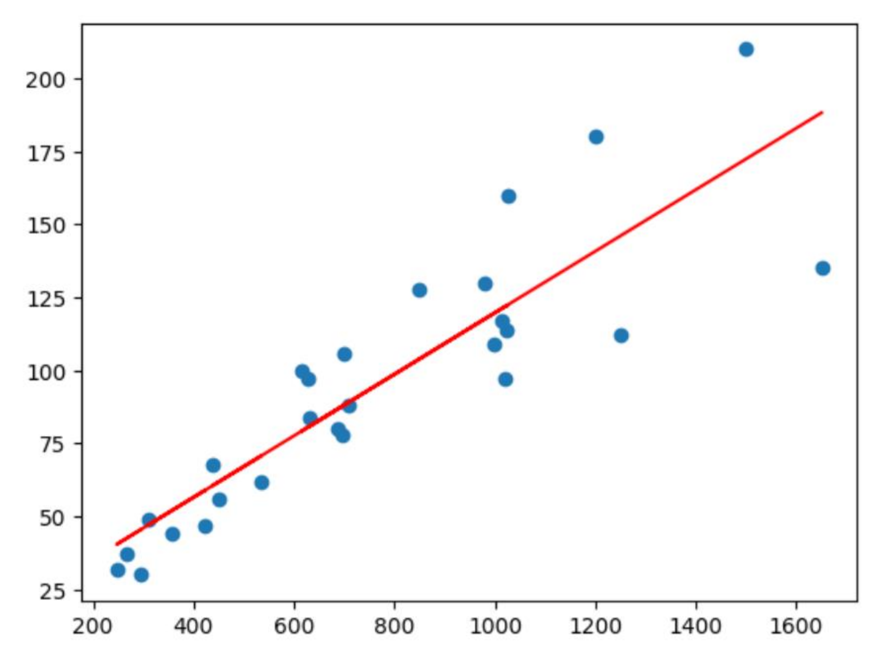
- This scenario can be represented as $Y = f(X) + \epsilon$, i.e. the target variable is a function of input variables X.
- Uncertainty always exists in the process of determining the value of the target variable and this motivates to represent this as a probability distribution $P(Y|X)$.

Linear Regression Model

- The linear regression model has the form $Y = \beta_0 + \sum_{j=1}^{d} \beta_j X_j$
- Here β_j are unknown parameter and X_j can come come from different sources:
- quantitative inputs
- transformation of quantitative inputs, such as log, square root or square
- basic expansion
- numeric or dummy coding of the qualitative inputs

Least Square Method

- The most popular estimation method is least squares, in which we pick the coefficients $\beta = \left(\beta_0, \beta_1, \ldots, \beta_p\right)^T$ to minimize the residual sum of squares.
- $RSS = \sum_{i=1}^{N} \left(y_j - f\left(x_j\right)\right)^2$
- $RSS = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{d} \beta_j x_{ij}\right)^2$



Simple Linear Regression with Intercept Only, $\beta_1 = 0$

- Linear regression with intercept only means $\beta_1 = 0$, the linear regression equation becomes
- $f(x_i) = \beta_0$
- Then the residual sum squares error becomes $RSS = \sum(y_i - \beta_0)^2$
- Take the first order derivative of RSS, and equalize it to zero, to obtain the value of β_0
- $\frac{\partial RSS}{\partial \beta_0} = \sum 2(y_i - \beta_0)(-1) = 0$
- $\beta_0 = \frac{1}{N}\sum y_i = \bar{y}$

Simple Linear Regression with slope only

- Linear regression with slope only means $\beta_0 = 0$, the linear regression equation becomes
- $f(x_i) = \beta_1 x_i$
- Then the residual sum squares error becomes $RSS = \sum(y_i - \beta_1 x_i)^2$
- Take the first order derivative of RSS, and equalize it to zero, to obtain the value of β_1
- $\frac{\partial RSS}{\partial \beta_0} = \sum 2(y_i - \beta_1 x_i)(-x_i) = 0$
- $\sum x_i y_i - \beta_1 \sum x_i^2 = 0 \Rightarrow \beta_1 = \frac{\sum x_i y_i}{\sum x_i^2}$

Simple Linear Regression with Slope and Intercept

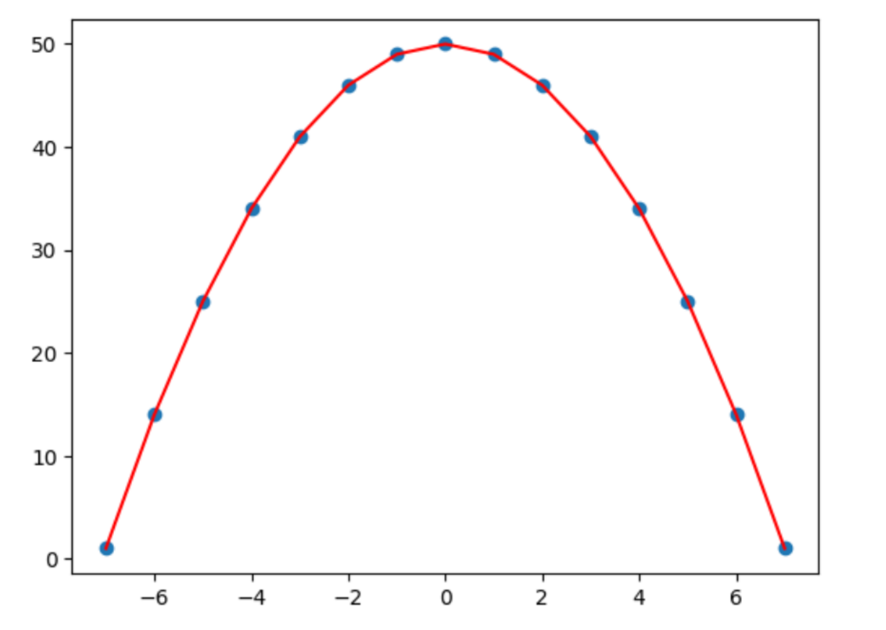
- The equation of linear regression is $f(x_i) = \beta_0 + \beta_1 x_i$
- The residual sum square error is $RSS = \sum(y_i - \beta_0 - \beta_1 x_i)$
- $\frac{\partial RSS}{\partial \beta_0} = \sum 2(y_i - \beta_0 - \beta_1 x_i)(-1)$
- $\frac{\partial RSS}{\partial \beta_1} = \sum 2(y_i - \beta_0 - \beta_1 x_i)(-x_i)$
- Equalize the first order derivative to zero, to obtain the parameters
- $\frac{\partial RSS}{\partial \beta_0} = 0 \Rightarrow \beta_0 = \frac{1}{N}(\sum y_i - \beta_1 \sum x_i)$

Simple Linear Regression with Slope and Intercept

- $\frac{\partial RSS}{\partial \beta_1} = 0 \Rightarrow \sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 = 0$
- $\beta_1 = \frac{\frac{1}{N}\sum x_i \sum y_i - \sum x_i y_i}{\frac{1}{N}(\sum x_i)^2 - \sum x_i^2}$
- The parameter values can be obtained using OLS method from the above equation.

Least Square Method

- $\mathcal{L} = (Y - X\beta)^T(Y - X\beta)$
- $\frac{\partial \mathcal{L}}{\partial \beta} = -2X^T(Y - X\beta)$
- $\frac{\partial^2 \mathcal{L}}{\partial \beta^2} = 2X^T X$
- Set the first order derivative to zero $\frac{\partial \mathcal{L}}{\partial \beta} = 0$
- $X^T(Y - X\beta) = 0 \Rightarrow \hat{\beta} = (X^T X)X^T Y$

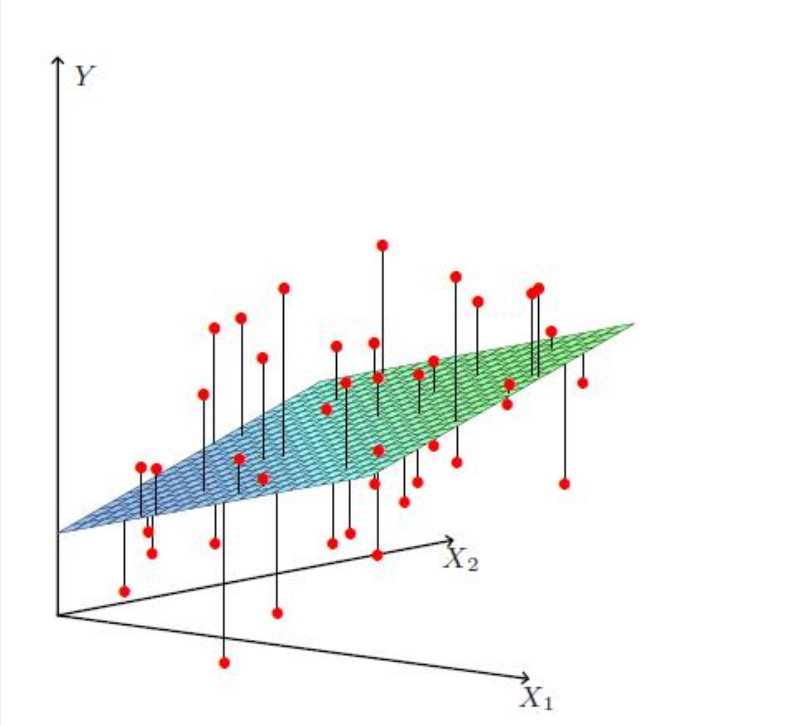


Multiple Regression

- The linear model with (d>1) inputs is called the multiple linear regression model. The least squares estimate for linear regression are best understood in terms of univariate model.
- Suppose we have single variable input with no intercept, that is
- $Y = X\beta + \epsilon$
- The least square estimate and residuals are
- $\hat{\beta} = \frac{\sum x_i y_i}{x_i^2}$
- $r = y_i - x_i\hat{\beta}$

Multiple Regression

- In vector notation $< x, y > = \sum x_i y_i = x^T y$
- Then the parameter and residual can be written as
- $\hat{\beta} = \frac{<x,y>}{<x,x>}$
- $r = y_i - x\hat{\beta}$

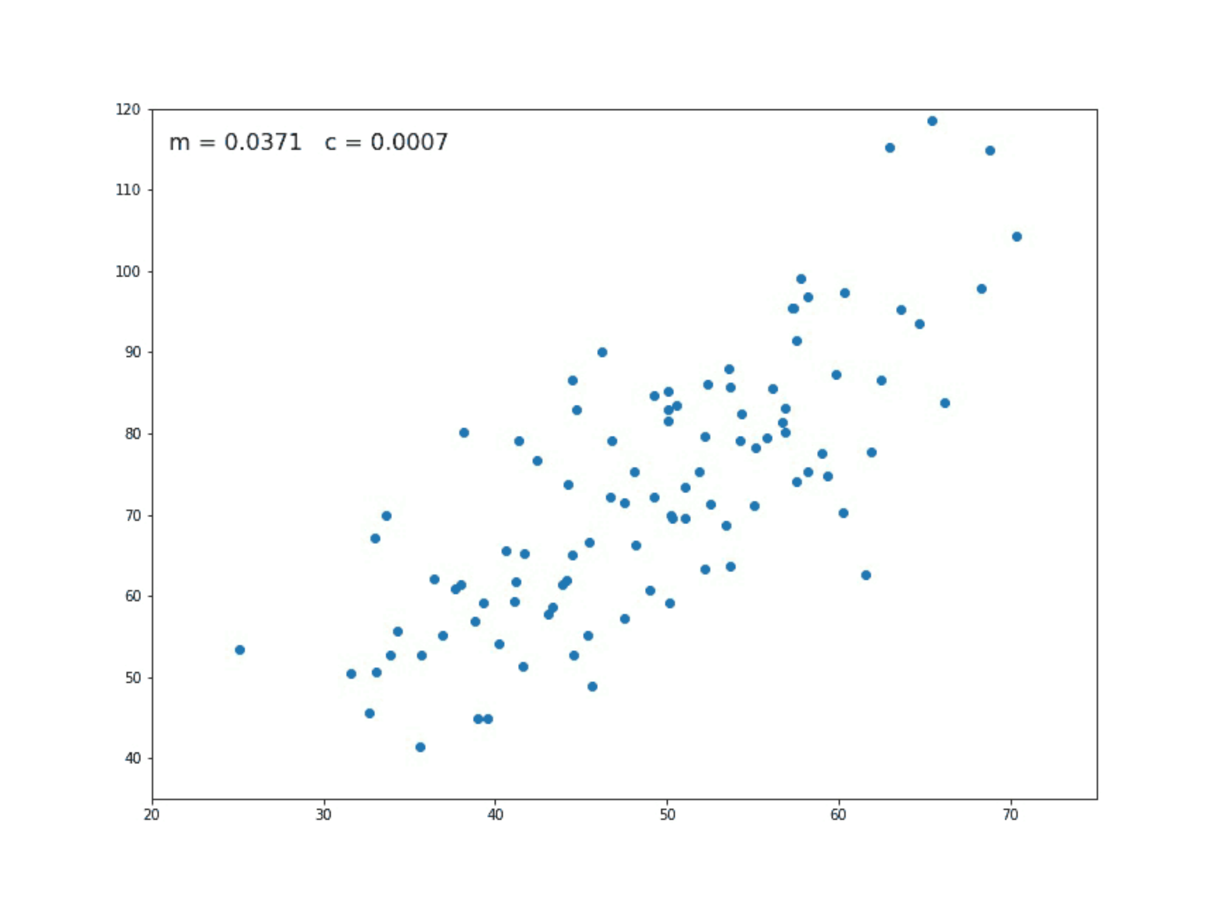


Gradient Descent Method

- In the equation of linear equation $y = \beta_0 + \beta_1 x$, β_0 is the intercept and β_1 is the slope.
- The objective is to find the values of β_0 and β_1, this will give best fit line with minimum error.
- Loss is the error, can be calculated using mean squared error function.
- $\mathcal{L} = \frac{1}{N}\sum(y_i - \beta_0 - \beta_1 x_i)^2$

Gradient Descent Algorithm

- The Gradient Descent is an iterative optimization algorithm to find the minimum of a function.



Gradient Descent Algorithm

1. Initialize the slope and intercept $\beta_0 = 0$ and $\beta_1 = 0$. Let $\mathcal{L}$ be the loss function and lr be the learning rate that controls the update of the parameters.
2. Calculate the partial derivative of the loss function with respect to the parameters.

$$\frac{\partial \mathcal{L}}{\partial \beta_0} = -\frac{2}{N}\sum(y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{\partial \mathcal{L}}{\partial \beta_1} = -\frac{2}{N}\sum(y_i - \beta_0 - \beta_1 x_i)(x_i)$$

Gradient Descent Algorithm

3. Update the value of the parameters using the following equation

$$\beta_0 = \beta_0 - lr \times \frac{\partial \mathcal{L}}{\partial \beta_0}$$

$$\beta_1 = \beta_1 - lr \times \frac{\partial \mathcal{L}}{\partial \beta_1}$$

4. Repeat this process until our loss is very small.