Linear Regression

Deborah Dormah Kanubala

April 12, 2020

Prerequisite for the video

Bit of Mathematics in here

Prerequisite for the video

Bit of Mathematics in here

1. Basic algebra

Prerequisite for the video

Bit of Mathematics in here

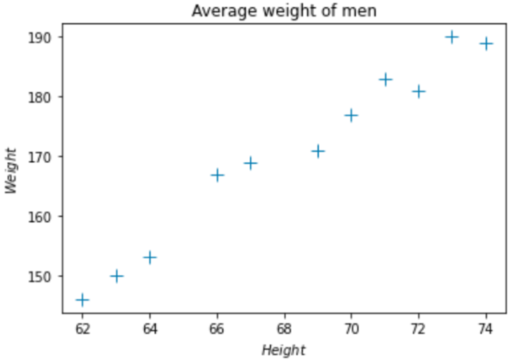
1. Basic algebra
2. Basic calculus

Overview

Suppose we have a data set giving a weight in pounds and height in inches;

Height(inches)	Weight(Pounds)
62	146
63	150
64	153
70	‘177
71	183
72	181
73	190
74	189
⋮	⋮

Plot



Source of data: https://www.cdc.gov/nchs/data/series/sr_11/sr11_014acc.pdf

Figure: Average weight of men between 18-79 years from 1960-1962 in US

Variables

1. **Independent:** Height X
2. **Dependent:** Weight y

Variables

1. **Independent:** Height X
2. **Dependent:** Weight y

Objective of Linear Regression

To determine the extent to which there is a linear relationship between a dependent variable and one or more independent variable.

Types of Linear Regression(LR)

1. **Simple LR:** Single independent variable used to predict value of dependent variable.

Types of Linear Regression(LR)

1. **Simple LR:** Single independent variable used to predict value of dependent variable.
2. **Multiple LR:** Two or more independent variables used to predict dependent variable

Types of Linear Regression(LR)

1. **Simple LR:** Single independent variable used to predict value of dependent variable.
2. **Multiple LR:** Two or more independent variables used to predict dependent variable

Our Task

To predict the value of the dependent variable based on independent variable. Hence, finding the best fit line.

Types of Linear Regression(LR)

1. **Simple LR:** Single independent variable used to predict value of dependent variable.
2. **Multiple LR:** Two or more independent variables used to predict dependent variable

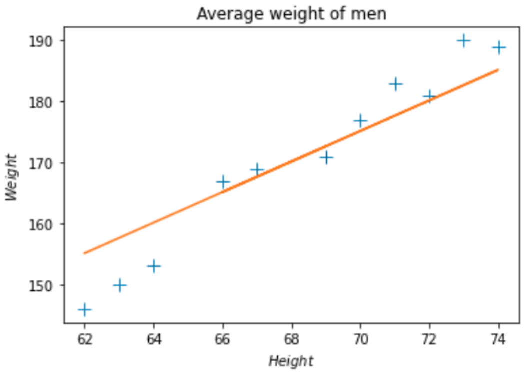
Our Task

To predict the value of the dependent variable based on independent variable. Hence, finding the best fit line.

Equation of a Line

$$y_i = \theta X_i + \epsilon_i$$

Best fit line



Source of data: `https://www.cdc.gov/nchs/data/series/sr_11/sr11_014acc.pdf`

Figure: Average weight of men between 18-79 years from 1960-1962 in US

Goal Linear Regression

$$\boxed{\ell(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta^T X_i)^2}$$

Goal Linear Regression

$$\boxed{\ell(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta^T X_i)^2}$$

How do we do this??

MLE

$$y_i = \theta^T X_i + \epsilon_i, \ \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

MLE

$$y_i = \theta^T X_i + \epsilon_i,\ \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

Make ϵ_i the subject

$$\epsilon_i = y_i - \theta^T X_i \tag{2}$$

MLE

$$y_i = \theta^T X_i + \epsilon_i,\ \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

Make ϵ_i the subject

$$\epsilon_i = y_i - \theta^T X_i \tag{2}$$

$$\mathbb{P}(\epsilon_i) = \mathbb{P}(y_i | X_i; \theta) \tag{3}$$

MLE

$$y_i = \theta^T X_i + \epsilon_i,\ \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

Make ϵ_i the subject

$$\epsilon_i = y_i - \theta^T X_i \tag{2}$$

$$\mathbb{P}(\epsilon_i) = \mathbb{P}(y_i | X_i; \theta) \tag{3}$$

$$\mathbb{P}(\epsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)} \tag{4}$$

MLE

$$y_i = \theta^T X_i + \epsilon_i, \; \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

Make ϵ_i the subject

$$\epsilon_i = y_i - \theta^T X_i \tag{2}$$

$$\mathbb{P}(\epsilon_i) = \mathbb{P}(y_i | X_i; \theta) \tag{3}$$

$$\mathbb{P}(\epsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)} \tag{4}$$

But $\mu = 0$; $x = y_i - \theta^T X_i$

MLE

$$\frac{1}{\sigma\sqrt{2\pi}}e^{\left(\frac{-(y_i-\theta^T X_i)^2}{2\sigma^2}\right)} \tag{5}$$

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2}\right)} \tag{5}$$

$\mathbb{P}(\epsilon_i)$ are i.i.d;

MLE

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2}\right)} \tag{5}$$

$\mathbb{P}(\epsilon_i)$ are i.i.d;

$$\mathbb{P}(y|X;\theta) = \prod_{i=1}^{n} \mathbb{P}(y|X;\theta) = \mathbb{L}(\theta) \tag{6}$$

MLE

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2}\right)} \tag{5}$$

$\mathbb{P}(\epsilon_i)$ are i.i.d;

$$\mathbb{P}(y|X;\theta) = \prod_{i=1}^{n} \mathbb{P}(y|X;\theta) = \mathbb{L}(\theta) \tag{6}$$

$$\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2}\right)} \tag{7}$$

MLE

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2}\right)} \tag{5}$$

$\mathbb{P}(\epsilon_i)$ are i.i.d;

$$\mathbb{P}(y|X;\theta) = \prod_{i=1}^{n} \mathbb{P}(y|X;\theta) = \mathbb{L}(\theta) \tag{6}$$

$$\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2}\right)} \tag{7}$$

Goal: $\max \mathbb{P}(y|X;\theta)$. How?

MLE

$$\frac{1}{\sigma\sqrt{2\pi}}e^{\left(\frac{-(y_i-\theta^T X_i)^2}{2\sigma^2}\right)} \tag{5}$$

$\mathbb{P}(\epsilon_i)$ are i.i.d;

$$\mathbb{P}(y|X;\theta) = \prod_{i=1}^{n}\mathbb{P}(y|X;\theta) = \mathbb{L}(\theta) \tag{6}$$

$$\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{\left(\frac{-(y_i-\theta^T X_i)^2}{2\sigma^2}\right)} \tag{7}$$

Goal: $\max \mathbb{P}(y|X;\theta)$. How?

MLE

$$\log \mathbb{L}(\theta) = \log \left[\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2} \right)} \right] \tag{8}$$

MLE

$$\log \mathbb{L}(\theta) = \log \left[\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2} \right)} \right] \tag{8}$$

Rule: $e^a . e^b = e^{a+b}$

$$\log \mathbb{L}(\theta) = \log \left[\left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2}} \right] \tag{9}$$

MLE

$$\log \mathbb{L}(\theta) = \log \left[\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2} \right)} \right] \tag{8}$$

Rule: $e^a . e^b = e^{a+b}$

$$\log \mathbb{L}(\theta) = \log \left[\left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{- \sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2}} \right] \tag{9}$$

Rule: $\log(ab) = log(a) + log(b)$

$$\log \mathbb{L}(\theta) = \log \left[\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(y_i - \theta^T X_i)^2}{2\sigma^2} \right)} \right] \tag{8}$$

Rule: $e^a.e^b = e^{a+b}$

$$\log \mathbb{L}(\theta) = \log \left[\left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2}} \right] \tag{9}$$

Rule: $\log(ab) = log(a) + log(b)$

$$\log \mathbb{L}(\theta) = \log \left[\left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \right] + \log \left[e^{-\sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2}} \right] \tag{10}$$

MLE

Rule: $\log(e^{u(x)}) = u(x)$ and $\log(a^n) = n \log a$

MLE

Rule: $\log(e^{u(x)}) = u(x)$ and $\log(a^n) = n \log a$

$$n \log \left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2} \tag{11}$$

MLE

Rule: $\log(e^{u(x)}) = u(x)$ and $\log(a^n) = n \log a$

$$n \log \left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2} \tag{11}$$

Interested in θ that maximizes this problem so we use the argmax

MLE

Rule: $\log(e^{u(x)}) = u(x)$ and $\log(a^n) = n \log a$

$$n \log \left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2} \tag{11}$$

Interested in θ that maximizes this problem so we use the argmax

$$\underset{\theta}{\operatorname{argmax}} \log \mathbb{L}(\theta) = \underset{\theta}{\operatorname{argmax}} \left(- \sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2} \right) \tag{12}$$

Rule: $\underset{\theta}{\operatorname{argmax}}\, f(\theta) = \underset{\theta}{\operatorname{argmax}}\, \alpha\, f(\theta)\ , \forall\, \alpha \in \mathbb{R}_+^*$

Rule: $\log(e^{u(x)}) = u(x)$ and $\log(a^n) = n \log a$

$$n \log \left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2} \tag{11}$$

Interested in θ that maximizes this problem so we use the argmax

$$\operatorname*{argmax}_{\theta} \log \mathbb{L}(\theta) = \operatorname*{argmax}_{\theta} \left(- \sum_{i=1}^{n} \frac{(y_i - \theta^T X_i)^2}{2\sigma^2} \right) \tag{12}$$

Rule: $\operatorname*{argmax}_{\theta} f(\theta) = \operatorname*{argmax}_{\theta} \alpha \, f(\theta) \,, \forall \, \alpha \, \in \, \mathbb{R}_{+}^{*}$

Note $\max f(\theta) = -\min(-f(\theta))$

This then produces the mean square error and minimizing this function is the interest.

$$\boxed{\ell(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta^T X_i)^2}$$