Gaussian Error Assumption Over Linear Regression

Nilotpal Sarkar

February 2021

1 Introduction

Linear regression is the methodology to model the relationship between two variables by fitting a line to the observed data. there are mainly two types of variables present which are independent and dependent variable.

2 Probabilistic Modelling

2.1 Linear Model

$$y_i \simeq \theta^T x_i$$

$$y_i = \theta^T x_i + \epsilon_i$$

where,

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Here ϵ_i's are i.i.d random variables.

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$
$$\Rightarrow p(\epsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{\epsilon_i^2}{2\sigma^2}\right)$$
$$\Rightarrow p(y_i - \theta^T x_i) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}\right)$$

However the conventional way is,

$$p(y_i|x_i;\theta) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}\right)$$

2.2 Parameter Estimation

Suppose, we are given with a dataset,

$$\mathcal{D} = \{x_i, yi\}_{i=1}^{m}$$

Then, Bayes' Theorem States that,

$$\begin{aligned} P(\theta|\mathcal{D}) &= P(\mathcal{D}|\theta).P(\theta) \\ &= P(\theta, \mathcal{D})\frac{1}{P(\mathcal{D})} \end{aligned} \tag{1}$$

2.3 Maximum Likelihood Estimation (MLE)

We will make a Gaussian model assumption with $\theta \in R^n$

$$\begin{aligned} \theta^* &= \underset{\theta}{\operatorname{argmax}}\ \mathcal{L}(\theta|\mathcal{D}) \\ &= \underset{\theta}{\operatorname{argmax}}\ P(\mathcal{D} \mid \ \theta) \\ &= \underset{\theta}{\operatorname{argmax}}\ P(y_1, x_1, ..., y_m, x_m; \theta) \\ &= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{m} P(y_i, x_i; \theta) \\ &= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{m} [P(y_i|x_i; \theta).P(x_i; \theta] \\ &= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{m} [P(y_i|x_i; \theta)].P(x_i) \\ &= \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{m} [P(y_i|x_i; \theta)] \\ &= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{m} logP(y_i|x_i; \theta) \\ &= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{m} [log(\frac{1}{\sqrt{2\pi}\sigma}) + log(exp(-\frac{(\theta^T x_i - y_i)^2}{2\sigma^2}))] \\ &= \underset{\theta}{\operatorname{argmax}}\ -\frac{1}{2\sigma^2} \sum_{i=1}^{m} (\theta^T x_i - y_i)^2 \\ &= \underset{\theta}{\operatorname{argmin}}\ \frac{1}{n} \sum_{i=1}^{m} (\theta^T x_i - y_i)^2 \end{aligned} \tag{2}$$

3 Conclusion

We thus come to the conclusion that the error assumption under linear assumption leads to least square assumption.