Under Gaussian assumption linear regression amounts to least square (ordinary least square)

Debojyoti Mondal

$25^{th} February, 2021$

Theorem 1. *prove that under Gaussian assumption linear regression amounts to least square*

Proof: Consider the linear model where target variables $_i$ and the inputs $_i$ are related via :

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

where ϵ_i is the error associated with the model. According to the theorem further assume that the ϵ_i are distributed IID (independently and identically distributed) according to a Gaussian distribution (also called a Normal distribution) with mean zero and some variance σ,i.e,

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

The density of ϵ_i is given by

$$p(\epsilon_i) \ = \ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\epsilon_i^2)}{2\sigma^2}\right)$$

This implies that

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

However, the conventional way to write the above probability is

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

which indicates that this is the distribution of y_i given x_i and parameterized by θ. The distribution of y_i is given as $y_i | x_i; \ \theta \ \sim \ \mathcal{N}(\theta^T x_i, \sigma^2)$
From Bayes Theorem:

$$p(\theta | D) \ = \frac{p(D|\theta) p(\theta)}{p(D)}$$

Where D is data, $p(\theta|D)$ is posterior probability, $p(\theta)$ is prior probability and $p(\theta|D)$ is data likelihood.

From Maximum Likelihood Estimation(MLE):

$$\begin{aligned}
\theta^* &= argmax_\theta \;\; L(\theta|D) \\
&= argmax_\theta \;\; p(\theta|D) \\
&= argmax_\theta \;\; p(y_1, x_1, y_2, x_2, ..., y_m, x_m; \theta) \\
&= argmax_\theta \;\; \prod_{i=1}^{m} \;\; p(y_i, x_i; \theta) p(x_i; \theta) \\
&= argmax_\theta \;\; \prod_{i=1}^{m} \;\; p(y_i, x_i; \theta) p(x_i) \\
&= argmax_\theta \;\; \prod_{i=1}^{m} \;\; p(y_i, x_i; \theta) \\
&= argmax_\theta \;\; \sum_{i=1}^{m} \;\; \log p(y_i, x_i; \theta) \\
&= argmax_\theta \;\; \sum_{i=1}^{m} \;\; \log \frac{1}{\sqrt{2\pi}\sigma} + \log \exp\left(\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}\right) \\
&= argmax_\theta \;\; \frac{-1}{2\sigma^2} \;\; \sum_{i=1}^{m} \;\; \left((y_i - \theta^T x_i)^2\right) \\
&= argmax_\theta \;\; \frac{-1}{m} \;\; \sum_{i=1}^{m} \;\; \left((y_i - \theta^T x_i)^2\right)
\end{aligned}$$

The above form represents the least square form.

Therefore it is proved that under Gaussian assumption linear regression amounts to least square (ordinary least square).