Assignment 1

Dipan Banik

26 February 2021

Introduction

One approach to solve Linear Regression Problem on a data set $D\{x_i, y_i\}_1^m$ is **Probabilistic Modelling**. In probabilistic models, we introduce a concept of random noise (ε_i) which measures the unknown error effects.

$$y_i = \theta^T X_i + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim N(0, \sigma^2)$$

[Note : ε_i's are independent and identically distributed as the observations are independent of each other]

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

However the conventional way to write the probability is,

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

Parameter Estimation :

The regression coefficients θ are generally estimated using Maximum Likelihood Estimation. In this process we try to estimate the parameters of the population by the observed data. $L(\theta|D)$ represents the likelihood of the parameter taking that value where D represents the data, and that likelihood is exactly same as the probability of getting that data when θ is the population parameter. So we shall find a vector θ such that $P(D|\theta)$ is maximized.

Estimating θ with MLE (Maximum Likelihood Estimator):

Theorem 1 *Under Gaussian assumption of the Noise in the prediction, Linear Regression amounts to least square i.e. ordinary least square.*

Proof:

$$
\begin{aligned}
\theta^* &= \arg\max_\theta L(\theta|D) \\
&= \arg\max_\theta P(D|\theta) \\
&= \arg\max_\theta P(y_1, x_1, y_2, x_2, \ldots, y_m, x_m; \theta) \\
&= \arg\max_\theta \prod_{i=1}^{m} P(y_i, x_i; \theta) && \text{[As, the data observations are independent]} \\
&= \arg\max_\theta \prod_{i=1}^{m} P(y_i|x_i; \theta) P(x_i|\theta) \\
&= \arg\max_\theta \prod_{i=1}^{m} P(y_i|x_i; \theta) && \text{[As, } x_i\text{'s are independent of } \theta\text{]} \\
&= \arg\max_\theta \sum_{i=1}^{m} [\log(\frac{1}{\sqrt{2\pi}\sigma}) + \log(\exp(\frac{-(y_i - \theta^T x_i)^2}{2\sigma^2}))] && \text{[As log is monotone increasing function]} \\
&= \arg\max_\theta -\frac{1}{2\sigma^2} \sum_{i=1}^{m} ((y_i - \theta^T x_i)^2) \\
&= \arg\min_\theta \frac{1}{m} \sum_{i=1}^{m} (y_i - \theta^T x_i)^2
\end{aligned}
$$

. . . which is nothing but the least square method. So, **Probabilistic Modelling under Gaussian assumption of noise is ultimately equivalent to Least square regression.**