Probabilistic modeling -Linear regression

Soumyadeep Banik

March 1, 2021

1 Linear Model:

In case of Linear regression we minimize the errors but here we are introducing a random variable ϵ_i [See (eq1)] chosen from Gaussian probabilistic distribution that is going to take care of all the errors or noises in the model. [?]

$$y_i \approx \theta^T x_i \dots \tag{1}$$

$\mathrm{y}_i = \theta^T x_i + \epsilon_i$ where $\epsilon_i\ N(0, \sigma^2)$

$\exists$ unique ϵ_i for each $y_i, x_i\ \forall i = 1, \dots, m$

And all ϵ_i come from a single probabilistic distribution. So, ϵ_i's are independent and identically distributed(iid) random variables.

Since, ϵ_i follows this probability distribution, we can write the following.

$$P(_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\epsilon_i^2}{2\sigma^2}} \Rightarrow P(y_i|x_i;\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}} \tag{2}$$

2 Parameter Estimation

Now, our task is to estimate the θ parameter from a given dataset
$D = X_i, Y_{i i=1}^{m}$

So, following the Bayes theorem, $P(\theta|D) = \frac{P(D|\theta) \cdot P(\theta)}{P(\theta)} = \frac{P(\theta, D)}{P(D)}$

Now goal is to choose those particular θ which maximizes the prob. $P(D|\theta)$.

So, we can treat this probability as a function of θ called $Likelihood\,Function$.

As we are choosing such θ that fit the given dataset, it's often called $Data\,Likelihood.$

3 Maximum Likelihood Estimation(MLE)

$\theta^* = \underset{\theta}{\text{argmax}}\ P(\theta|D)$

$Since$ $\mathrm{y}_i = x_i + \epsilon_i$ and each $\epsilon_i' s$ are independent, we can write

$\underset{\theta}{\text{argmax}}\ P(\theta|D) = \underset{\theta}{\text{argmax}}\, P(Y_1, X_1, Y_2, X_2, \ldots, Y_m, X_m; \theta)$
$= \underset{\theta}{\text{argmax}}\ \prod_{i=1}^{m} P(Y_i, X_i; \theta) \qquad\qquad as P(A,B) = P(A|B) \cdot P(B)$
$= \underset{\theta}{\text{argmax}}\ \prod_{i=1}^{m} P(Y_i|X_i; \theta)\, P(X_i; \theta)$
$= \underset{\theta}{\text{argmax}}\ \prod_{i=1}^{m} P(Y_i|X_i; \theta)\, P(X_i)$ $\qquad$ X$_i$'s are independent of θ
$=\underset{\theta}{\text{argmax}}\ \prod_{i=1}^{m} P(Y_i|X_i; \theta)$ $\qquad\qquad$ since we are interested only in θ,
$=\underset{\theta}{\text{argmax}}\ \sum_{i=1}^{m} \log P(Y_i|X_i; \theta)$
$= \underset{\theta}{\text{argmax}}\ \sum_{i=1}^{m} [\log \frac{1}{\sigma\sqrt{2\pi}} + \log e^{-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}}]$
Since, the log term is indepedent of θ, we can eliminate it
$= \underset{\theta}{\text{argmax}}\ -\frac{1}{2\sigma^2} \sum_{i=1}^{m} (\theta^T x_i - y_i)^2$
$= \underset{\theta}{\text{argmin}}\ \frac{1}{m} \sum_{i=1}^{m} (\theta^T x_i^2 - y_i)$

Therefore, under Gaussian assumption linear regression amounts to least square.