Assignment

aritraraut

February 2021

Statement : under Gaussian assumption linear regression amounts to least square (ordinary least square).
Proof: Let us assume that the target variables and the inputs are related as

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

Where, ϵ_i's are random noise to model unknown effects and they are IID(independently and identically distriubuted) random variables.

$$\epsilon_i \sim N(0, \sigma^2),\ [\sigma = standard\ deviation\ of\ normal\ distribution]$$

So, we know,

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

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

However, the conventional way to write the probability is,

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

the notation indicates that this is the distribution of x_i and y_i and parameterized by θ.The probability of the data is given by $p(\bar{y}|X;\theta)$.The quantity is typically viewed as a function of $\bar{y}$,for fixed value of θ.when we wish to explicitly see this as function of θ,we call it the likelihood function.

$$L(\theta) = L(\theta, x, \bar{y}) = p(\bar{y}|X;\theta)$$

$$= \prod_{i=1}^{m} p(y_i|x_i;\theta)$$

$$= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2})$$

Now given this probabilistic model relating the y_i's and the x_i's,what is a reasonable way of choosing our best guess of parameters θ?.The principal of **maximum likelihood** says that we should choose θ so as to make the data as high probability as possible.So, we should maximize L(θ).

In perticular,the derivations will be a bit simpler if we maximize the log likelihood:

$$\begin{aligned}
\ell(\theta) &= logL(\theta) \\
&= log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}) \\
&= \sum_{i=1}^{m} log(\frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2})) \\
&= m\ log\frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2}.\frac{1}{2} \sum_{i=1}^{m} (y_i - \theta^T x_i)^2 \\
&= -\frac{1}{2\sigma^2} \sum_{i=1}^{m} (y_i - \theta^T x_i)^2
\end{aligned}$$

Hence,under gaussian error assumption MLE with parameter θ,is same as least square minimization.