AI511

MT2021130

⇒ Task: Derive Maximum likelihood function for normal/gaussian distribution.

Preamble:

Data $\{x_i\}_{i=1:N}$

Gaussian model

$$p(x_i | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Maximum Likelihood function,

$$p(X|\theta) = p(x_1, x_2 \dots x_N|\theta)$$

⇒ Proof:

Here, we're assuming that $x_1, x_2, \dots x_n$ are independent

$$\text{so } p(X|\theta) = p(x_1|\theta)\, p(x_2|\theta) \dots p(x_n|\theta)$$

$$= \prod_{i=1}^{n} p(x_i|\theta)$$

for gaussian $\theta = \mu, \sigma$

$$P(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$

taking log on both side

$$\ln(P(x|\theta)) = \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}\right) \cdots \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\right) \quad (1)$$

⇒ to simplify process just taking 'ln' of first term then put in result for all

$$\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}\right)$$

$$= -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_1-\mu)^2}{2\sigma^2}$$

$$= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{(x_1-\mu)^2}{2\sigma^2}$$

$$= -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{(x_1-\mu)^2}{2\sigma^2} \quad (2)$$

putting eq (2) result into eq (1) for all the terms we get

$$\ln(P(x|\theta)) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \left[\frac{(x_1-\mu)^2}{2\sigma^2} + \dots \frac{(x_n-\mu)^2}{2\sigma^2}\right]$$

now, to minimize MLF we differentiate w.r.t θ which in case of normal distribution is μ, σ.

so, $$\frac{\partial \ln p(x|\theta)}{\partial \mu} = 0 - 0 + \frac{(x_1-\mu)}{\sigma^2} + \dots \frac{(x_n-\mu)}{\sigma^2}$$

$$= \frac{1}{\sigma^2}\left[(x_1 + \dots + x_n) - n\mu\right] \quad (3)$$

for, $$\frac{\partial \ln p(x|\theta)}{\partial \sigma} = -\frac{n}{\sigma} + \left[\frac{(x_1-\mu)^2}{\sigma^3} + \dots + \frac{(x_n-\mu)^2}{\sigma^3}\right]$$

$$= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left[(x_1-\mu)^2 + \dots + (x_n-\mu)^2\right] \quad (4)$$

now, equating eq (4) w.r.t. 0

$$\frac{1}{\sigma^2}\left[(x_1+x_2+\ldots+x_n)-n\mu\right]=0$$

$$(x_1+x_2+\ldots+x_n)-n\mu=0$$

$$\boxed{\mu=\frac{x_1+\ldots+x_n}{n}}$$

similarly, for eq (5)

$$-\frac{n}{\sigma}+\frac{1}{\sigma^3}\left[(x_1-\mu)^2+\ldots+(x_n-\mu)^2\right]=0$$

$$n=\frac{1}{\sigma^2}\left[(x_1-\mu)^2+\ldots+(x_n-\mu^2)^2\right]=0$$

$$\sigma^2=\frac{1}{n}\left[(x_1-\mu)^2+\ldots+(x_n-\mu)^2\right]=0$$

$$\boxed{\sigma=\sqrt{\frac{(x_1-\mu)^2+\ldots+(x_n-\mu)^2}{n}}}$$

so, optimal parameters for normal distribution using MLE ~~is~~ are

$$\mu=\frac{\sum_{i=1}^{n} x_i}{n}$$

$$\sigma=\sum_{i=1}^{n}\sqrt{\frac{(x_i-\mu)^2}{n}}$$