# Machine Learning

Dr. Mehran Safayani

safayani@iut.ac.ir

safayani.iut.ac.ir

https://www.aparat.com/mehran.safayani

https://github.com/safayani/machine_learning_course

Department of Electrical and computer engineering, Isfahan university of technology, Isfahan, Iran

# MLE (Maximum Likelihood)

## یادآوری

توزیع گاوسی:

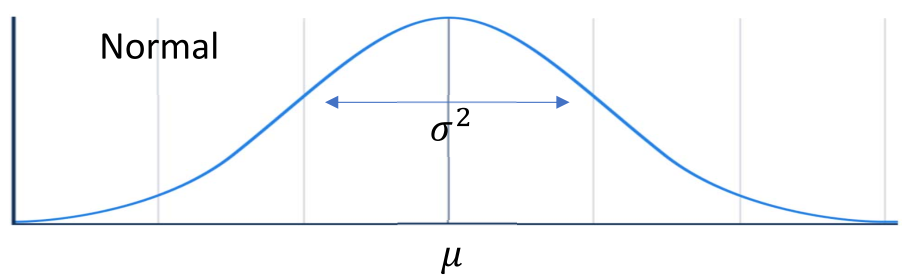
$$P(y \mid \mu , \sigma^2) = N(y \mid \mu , \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)$$

اگر y به صورت بردار باشد دارای میانگین $\vec{\mu}$ و کوواریانس $\Sigma$ است:

$$N(y \mid \mu , \Sigma) = \frac{1}{\sqrt{(4\pi)^D \det(\Sigma)}} \exp\left(\frac{-1}{2} (y-\mu)^T \Sigma^{-1} (y-\mu)\right)$$

x , y دو متغیر تصادفی مستقل هستند اگر:

$$P(x , y) = P(x)\, P(y)$$



می خواهیم تابع چگالی $P(x)$ را به دست آوریم.

یک تابع چگالی احتمالی پارامتری برای $P(x)$ تعریف میکنیم.

با فرض وجود مشاهده های $X = (x^1, \dots, x^n)$ سعی می کنیم پارامترهای مدل را بهینه کنیم تا احتمال $P(x)$ بیشینه گردد.

$$X = (x^1, \dots, x^n)$$

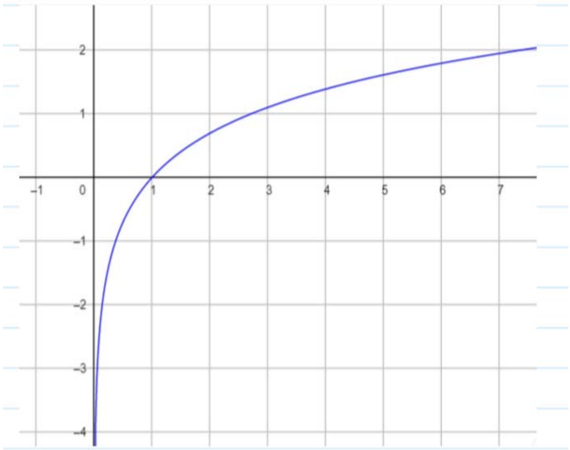
فرض می کنیم هر داده از توزیع $P(x \mid \theta)$ به صورت مستقل به دست آمده اند.

$$P(X \mid \theta) = \prod_{n=1}^{N} P(x^n \mid \theta) = L(\theta)$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, L(\theta)$$

$$\operatorname{Log} P(X \mid \theta) = \sum_{n=1}^{N} \log P(x^n \mid \theta) = \log L(\theta)$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, L(\theta) = \underset{\theta}{\operatorname{argmax}}\, \log L(\theta)$$



Let $x_1, x_2, \ldots, x_n$ be a random sample from a normal distribution with unknown mean $\mu$ and variance $\sigma^2$.

Find Maximum Likelihood estimators of mean $\mu$ and variance $\sigma^2$.

# Answer

In finding the estimators , the first thing we will do is write the probability density function as a function of $\theta_1 = \mu$ and $\theta_2 = \sigma^2$

$$f(x_i \,;\, \theta_1 \,,\, \theta_2) = \frac{1}{\sqrt{\theta_2}\,\sqrt{2\pi}} \exp\left[\frac{-(x_i-\theta_1)^2}{2\theta_2}\right]$$

For $-\infty < \theta_1 < \infty$ and $0 < \theta_2 < \infty$. We do this so as not to cause confusion when taking the derivative of the likelihood with respect to $\sigma^2$. Now , that makes the likelihood function:

$$L(\theta_1 \,,\, \theta_2) = \prod_{i=1}^{n} f(x_i \,;\, \theta_1 \,,\, \theta_2) = \theta_2^{-n/2}\,(2\pi)^{-n/2} \exp\left[\frac{-1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2\right]$$

And therefore the log of the likelihood function:

$$\text{Log L}(\theta_1 , \theta_2) = \frac{-n}{2} \log \theta_2 - \frac{n}{2} \log (2\pi) - \frac{\sum(x_i - \theta_1)^2}{2\theta_2}$$

Now, upon taking the partial derivative of the log likelihood with respect to $\theta_1$ , and setting to 0 , we see that a few things cancel each other out , leaving us with:

$$\frac{\partial \text{ Log L}(\theta_1 , \theta_2)}{\partial \theta_1} = \frac{-2 \sum(x_i - \theta_1)(-1)}{2\theta_2} \equiv 0$$

Now, multiplying through by $\theta_2$ and distributing the summation , we get:

$$\sum(x_i - n\theta_1) = 0$$

Now , solving for $\theta_1$ and putting on its hat we have shown that the maximum likelihood estimate of $\theta_1$ is :

$$\hat{\theta}_1 = \hat{\mu} = \frac{\sum x_i}{n} = \bar{x}$$

Now for $\theta_2$ taking the partial derivative of the log likelihood with respect to $\theta_2$ , and setting to 0 , we get:

$$\frac{\partial \text{ Log L}(\theta_1\ ,\ \theta_2)}{\partial \theta_2} = \frac{-n}{2\theta_2} + \frac{\sum(x_i - \theta_1)^2}{2\theta_2^2} = 0$$

Multiplying through by $2\theta_2^2$:

$$\frac{\partial \text{ Log L}(\theta_1\ ,\ \theta_2)}{\partial \theta_2} = [\frac{-n}{2\theta_2} + \frac{\sum(x_i - \theta_1)^2}{2\theta_2^2} = 0] * 2\theta_2^2$$

We get:

$$-n\theta_2 + \sum(x_i - \theta_1)^2 = 0$$

And , solving for $\theta_2$ , and putting on its hat , we have shown that the maximum likelihood estimate of $\theta_2$ is:

$$\hat{\theta}_2 = \hat{\sigma}^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$