machine learning prerequisites workshop
probability and statistics

Hamid R Rabiee – Zahra Dehghanian

Spring 2025

Sharif University of Technology

فهرست مطالب

فضای نمونه (Sample Space)

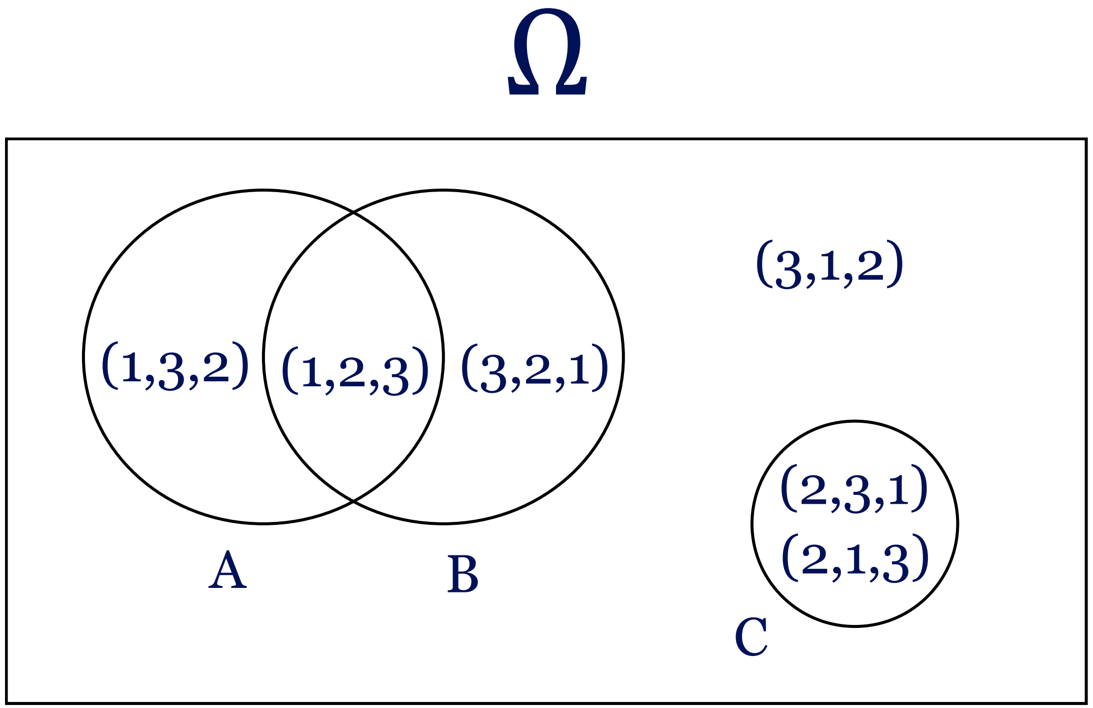
(شماره شرکت کننده سوم , شماره شرکت کننده دوم , شماره شرکت کننده اول)

Ω

(1,2,3)		(2,3,1)
	(1,3,2)	
(3,2,1)		
(3,1,2)		(2,1,3)

$$\Omega = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}$$

پیشامد (Event)



$A = \{(1,2,3), (1,3,2)\}$

$B = \{(1,2,3), (3,2,1)\}$

$C = \{(2,3,1), (2,1,3)\}$

اصول احتمال (Probability Axioms)

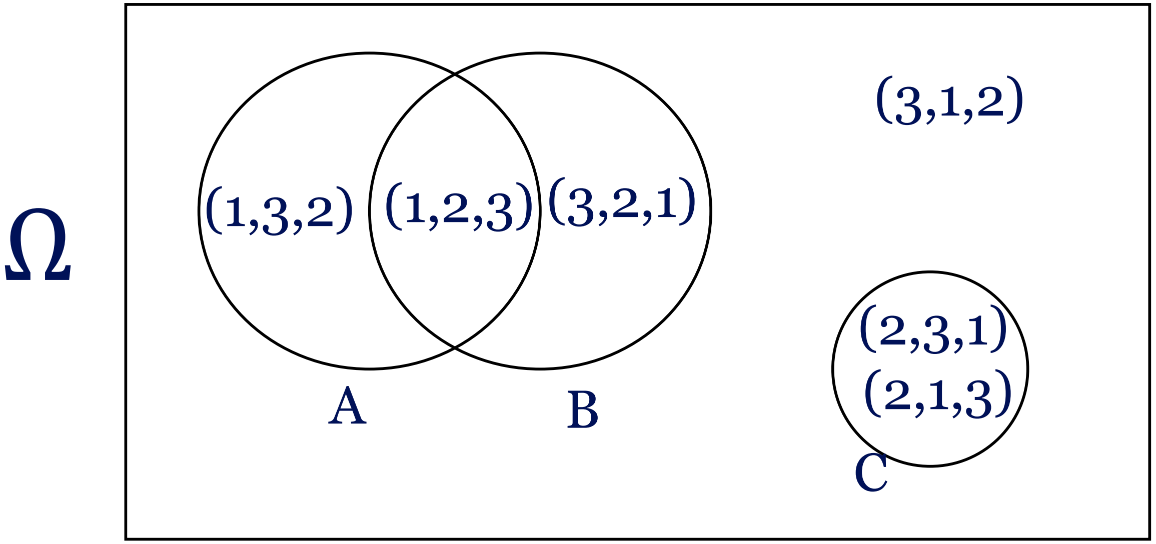
۱) به ازای هر پیشامد داریم : $0 \leq P(A) \leq 1$

۲) $P(\Omega) = 1$

۳) برای پیشامدهای ناسازگار داریم : $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

احتمال شرطی (Conditional Probability)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



$$P(A|B) = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2}$$

پیشامدهای مستقل (Independent Events)

$$P(A \cap B) = P(A)P(B)$$

قضیه بیز (Bayes Theorem)

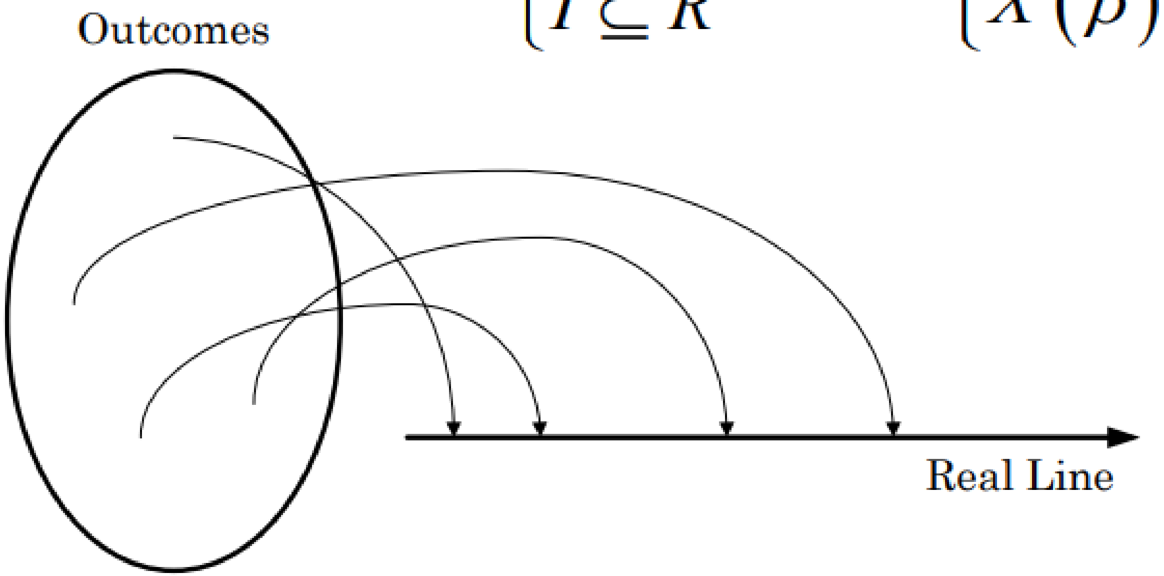
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

قانون احتمال کل (Law of total probability)

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

متغیر تصادفی (Random Variable)

$$\begin{cases} F \subseteq P(\Omega) \\ I \subseteq R \end{cases} : \begin{cases} X : F \to I \\ X(\beta) = r \end{cases}$$



۱) پیوسته : تابع چگالی احتمال

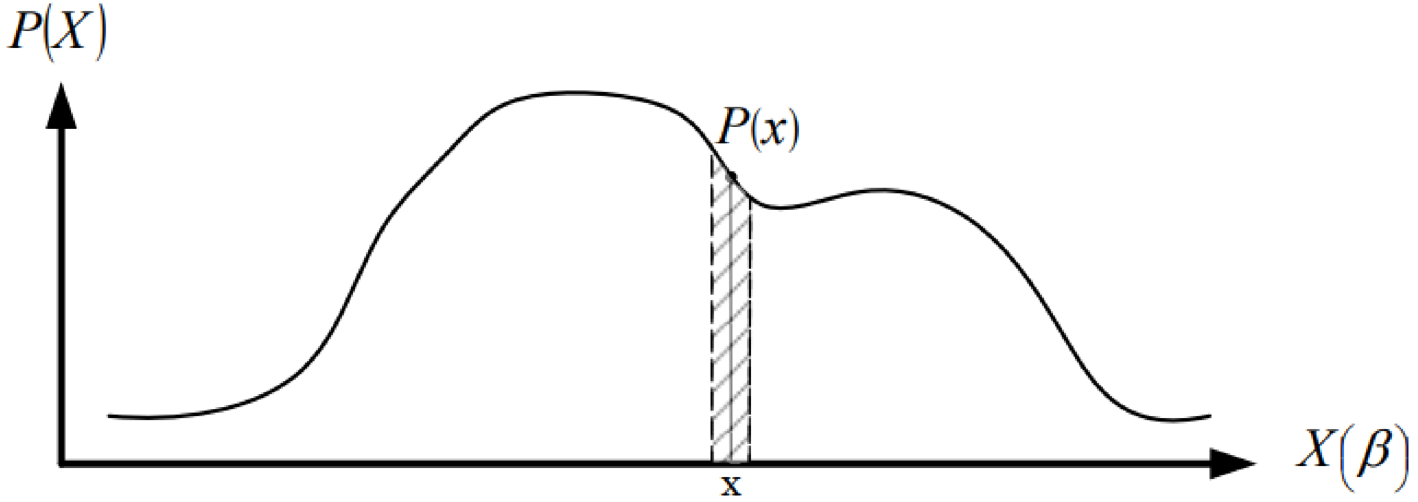
۲) گسسته : تابع جرم احتمال

تابع جرم احتمال (probability mass function - PMF)

$$f_X(x) = P(X = x)$$

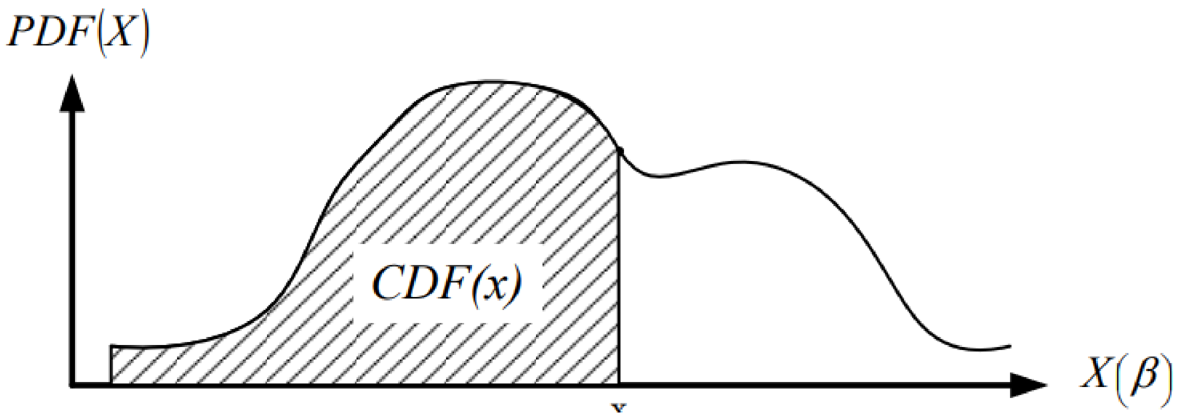
تابع چگالی احتمال (probability density function - PDF)

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$



تابع توزیع تجمعی (cumulative distribution function - CDF)

$$F_X(x) = P(X \leq x)$$



$$F_X(x) = \begin{cases} F_X(x) = \sum_{X \leq x} f_X(x) \\ F_X(x) = \int_{-\infty}^{x} f_X(x) \quad \Longrightarrow \frac{d}{dx} F_X(x) = f_X(x) \end{cases}$$

امید ریاضی (Expected Value)

$$E[X] = \mu = \begin{cases} \sum_{x \in \Omega} x f_X(x) \\ \\ \int_{-\infty}^{\infty} x f_X(x) \end{cases}$$

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

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

Law of the unconscious statistician - LOTUS

$$E[g(X)] = \begin{cases} \sum_{x \in \Omega} g(x) f_X(x) \\ \int_{-\infty}^{\infty} g(x) f_X(x) \end{cases}$$

واریانس (Variance)

$$Var(X) = \sigma^2 = E[(X-\mu)^2] = E[X^2] - E[X]^2$$

$$\mu = E[X]$$

$$Var(X) = \begin{cases} \sum_{x\in\Omega}(X-\mu)^2 f_X(x) \\ \\ \int_{-\infty}^{\infty}(x-\mu)^2 f_X(x) \end{cases}$$

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

$$Var(aX+b) = a^2Var(X)$$

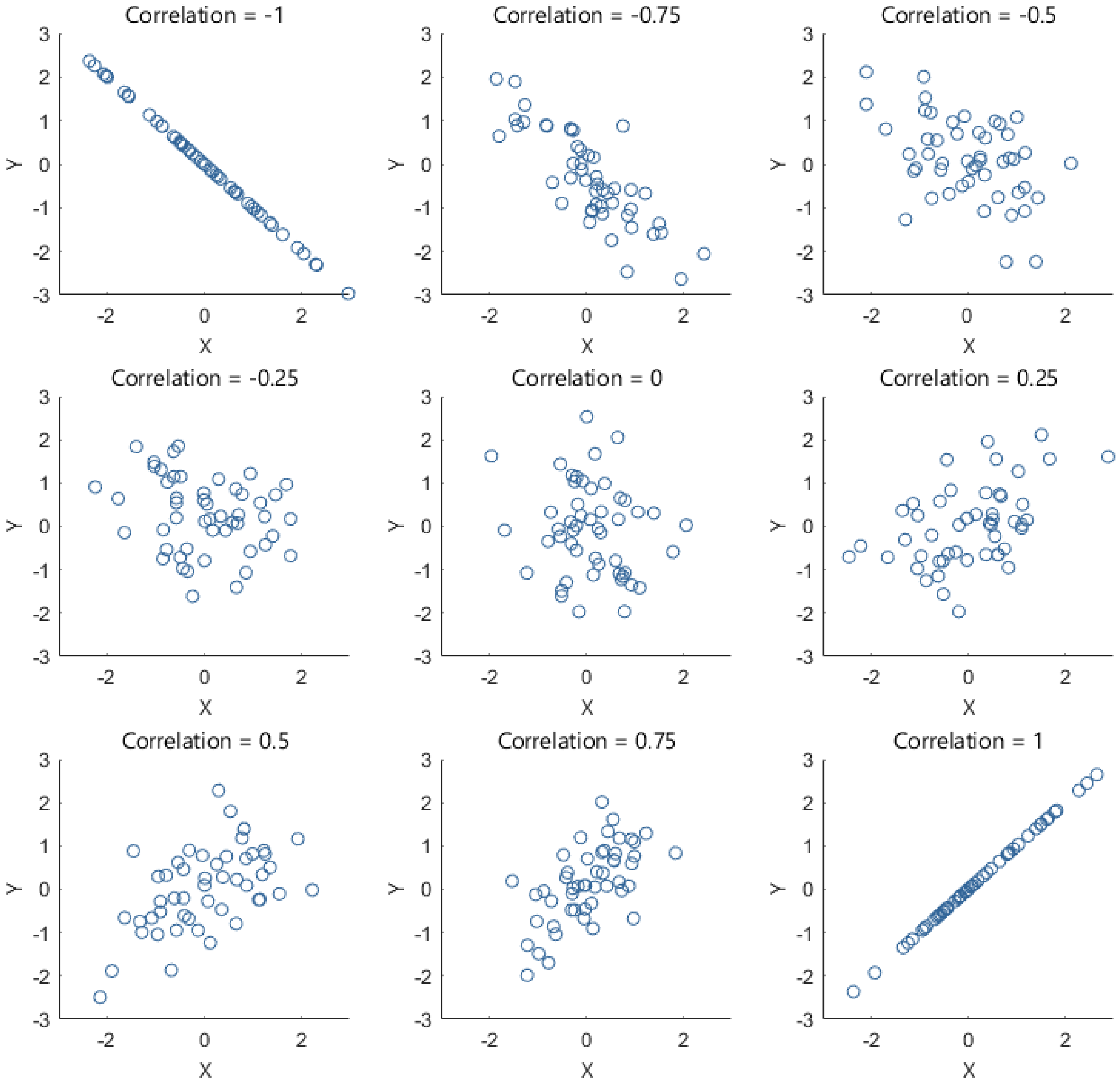
کوواریانس (Covariance)

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

$$Cov(X,Y) = \frac{\sum\left((x_i - \bar{x})(y_i - \bar{y})\right)}{n - 1}$$

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

همبستگی (correlation)

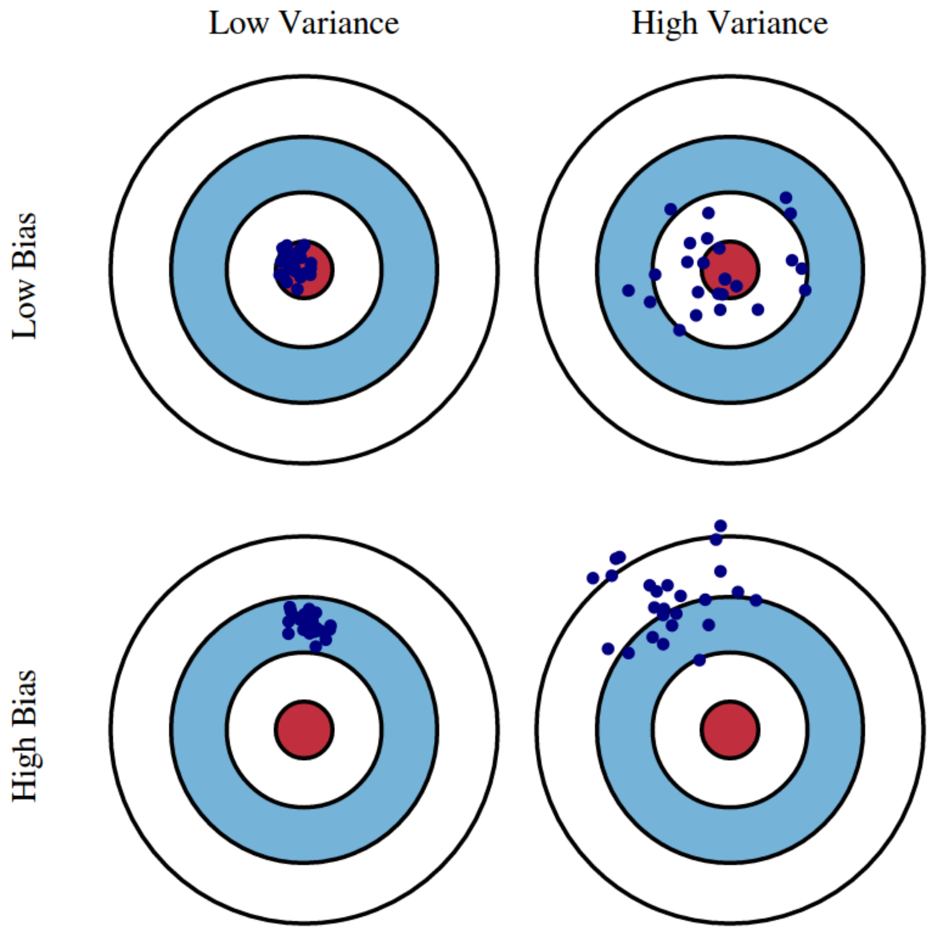


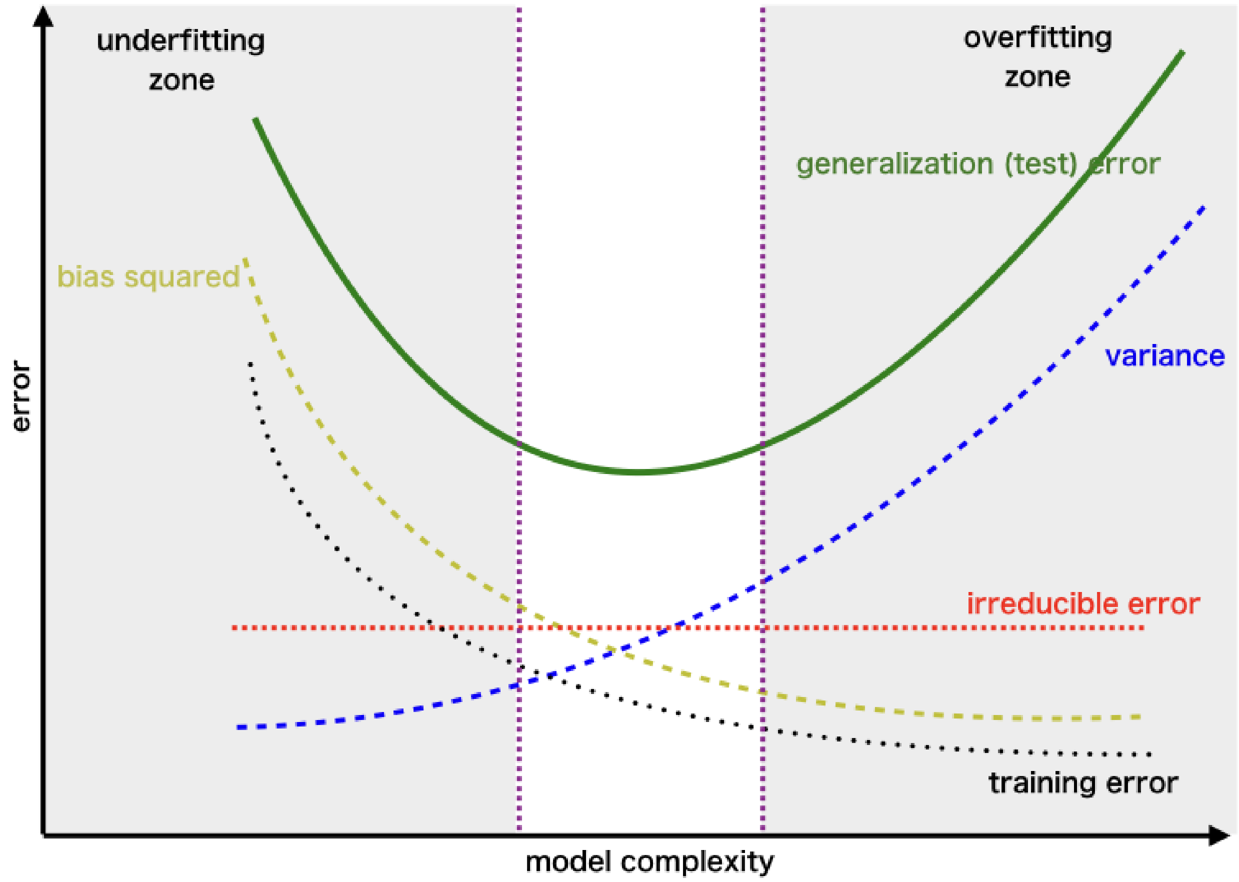
$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

اریب (Bias)

$$Bias(\hat{\theta}) = E[\hat{\theta} - \theta]$$

بایاس–واریانس تریدآف (bias-variance tradeoff)





متغیر تصادفی برنولی (Bernoulli)

$$X \sim Br(p)$$

$$P(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & otherwise \end{cases}$$

$$f_X(x) = p^x (1 - p)^{1-x}$$

$$E[X] = p$$

$$Var(X) = p(1 - p)$$

متغیر تصادفی دوجمله‌ای (Binomial)

$$X \sim Bin(n, p)$$

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & 0 \le x \le n \\ 0 & otherwise \end{cases}$$

$$E[X] = np$$

$$Var(X) = np(1-p)$$

متغیر تصادفی یکنواخت (Uniform Distribution)

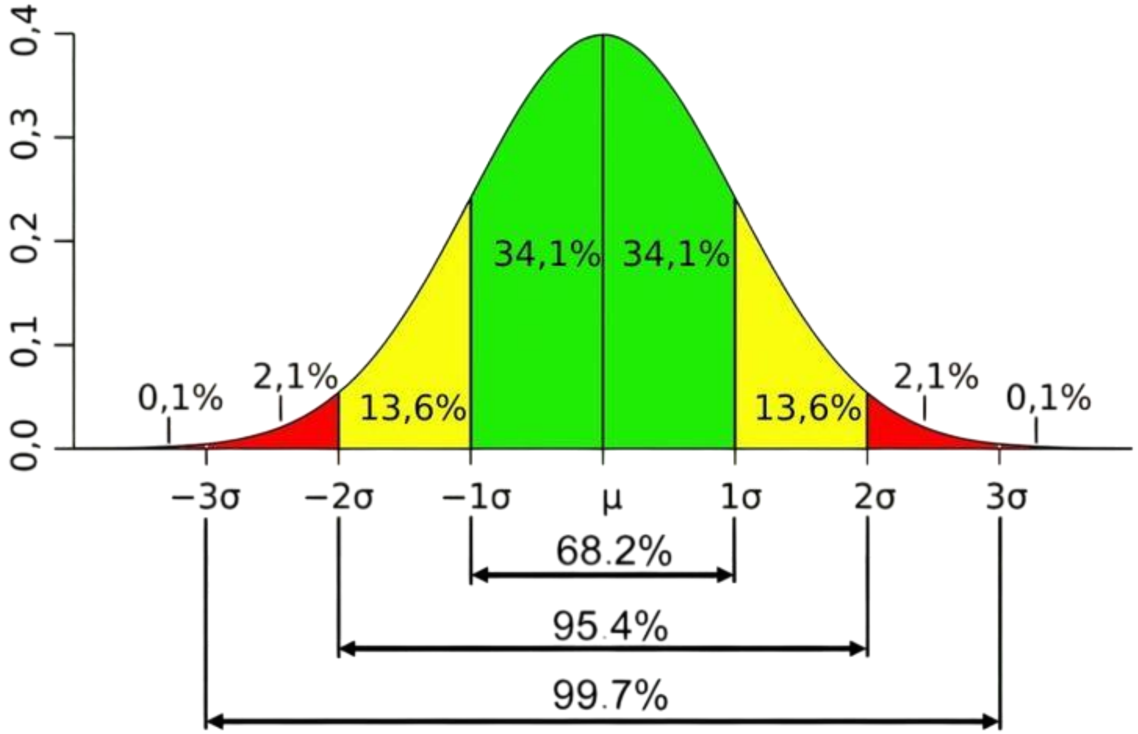
$$X \sim U(a, b)$$

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases}$$

$$E[X] = \frac{a+b}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

توزیع نرمال (Normal Distribution)



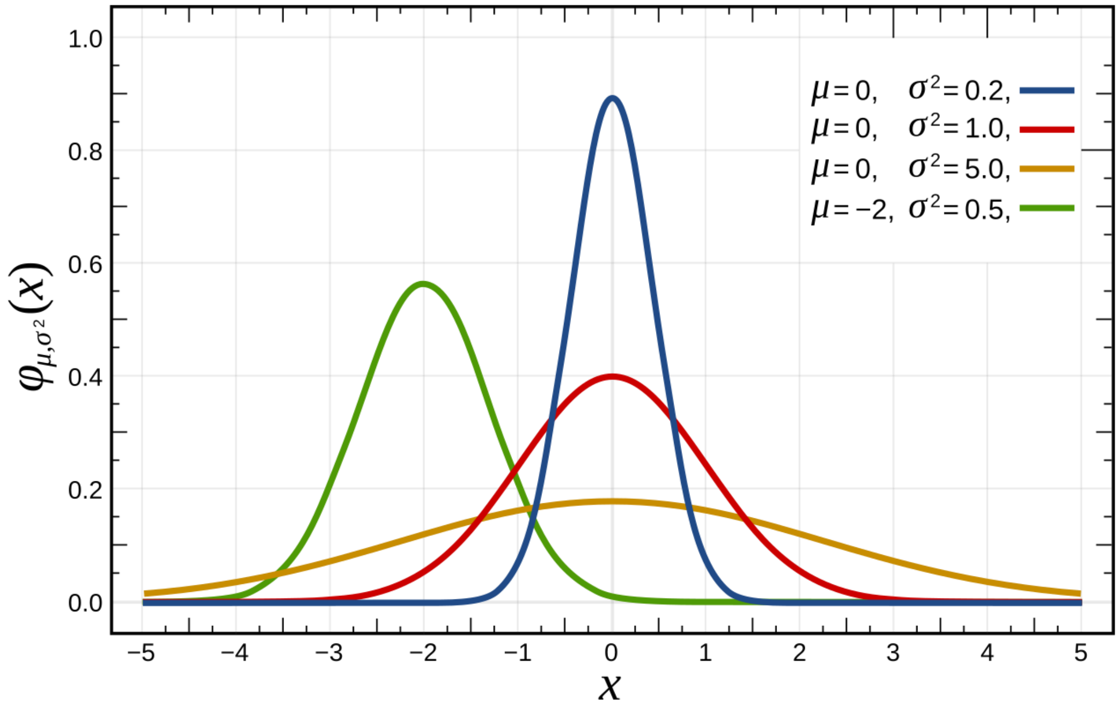
$$X \sim N(\mu, \sigma^2)$$

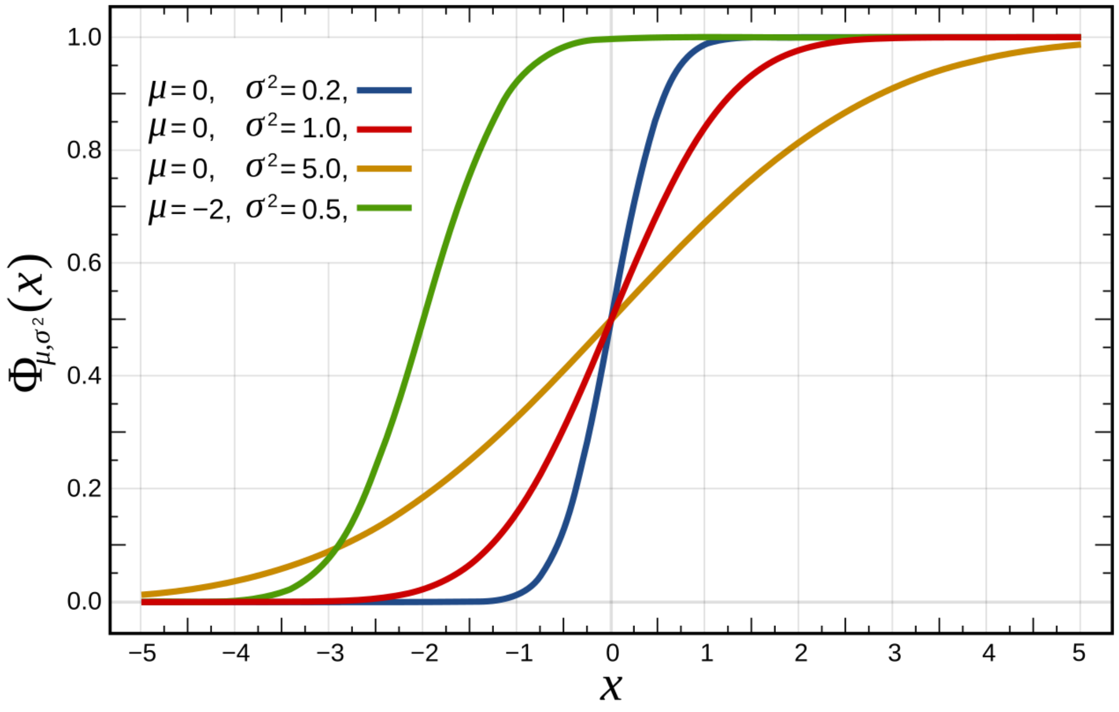
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$E[X] = \mu$$

$$Var(X) = \sigma^2$$

توزیع نرمال (Normal Distribution)





توزیع احتمال توام (Joint probability distribution)

$$F_{X,Y}(x,y) = P(X \le x \,, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(x,y)\,dy\,dx$$

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

قضیه حد مرکزی (Central Limit Theorem)

اگر $X_1, X_2, \ldots, X_n$ مستقل و دارای توزیع یکسان (idd) باشند به طوری که $E[X_i] = \mu$ و $Var[X_i] = \sigma^2$ ، آنگاه :

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

* idd : Independent and Identically Distributed

قانون اعداد بزرگ (law of large numbers)

اگر $X_1, X_2, \ldots, X_n$ مستقل و دارای توزیع یکسان (idd) باشند به طوری که $E[X_i] = \mu$، به ازای هر $\varepsilon > 0$

داریم :

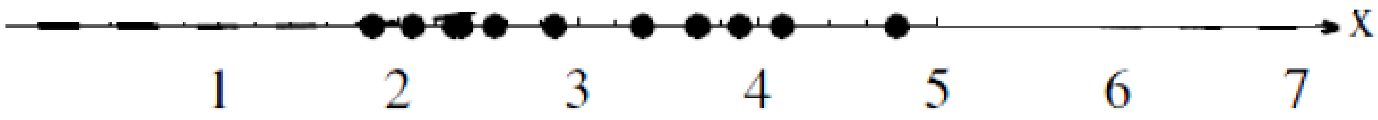
$$P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| < \varepsilon\right\} \to 0 \quad \text{اگر} \quad n \to \infty$$

یا به صورت دیگر :

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu$$

* idd : Independent and Identically Distributed

Maximum Likelihood Estimation (MLE)



$$P(x|\mu) = N(x|\mu, 1)$$

Likelihood, Posterior and Prior

$$P(\theta|Y) = \frac{P(Y|\theta)\ P(\theta)}{P(Y)}$$

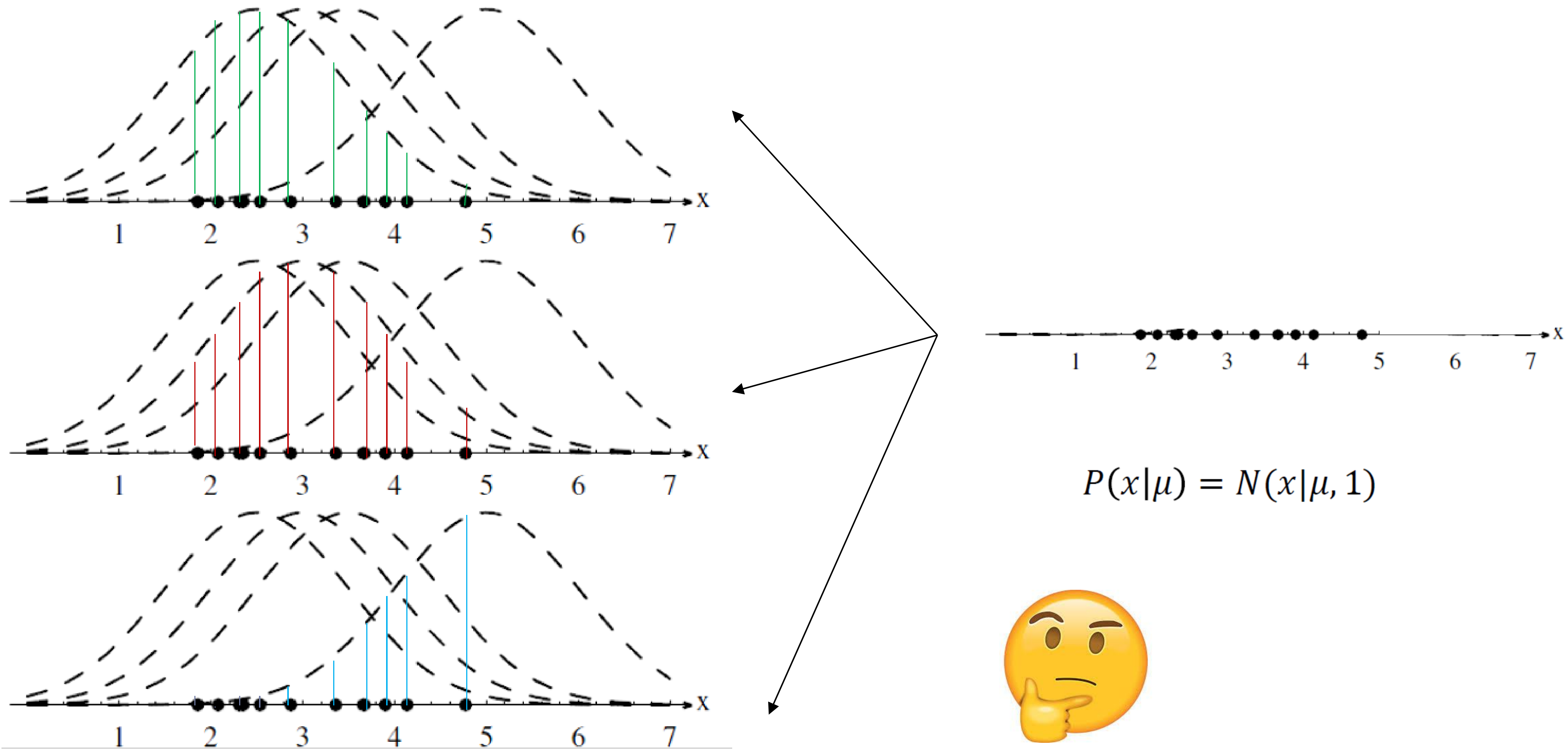
$$\propto P(Y|\theta)\ P(\theta)$$

Posterior: $P(\theta|Y)$

Likelihood: $P(Y|\theta)$

Prior: $P(\theta)$

Maximum Likelihood Estimation (MLE)



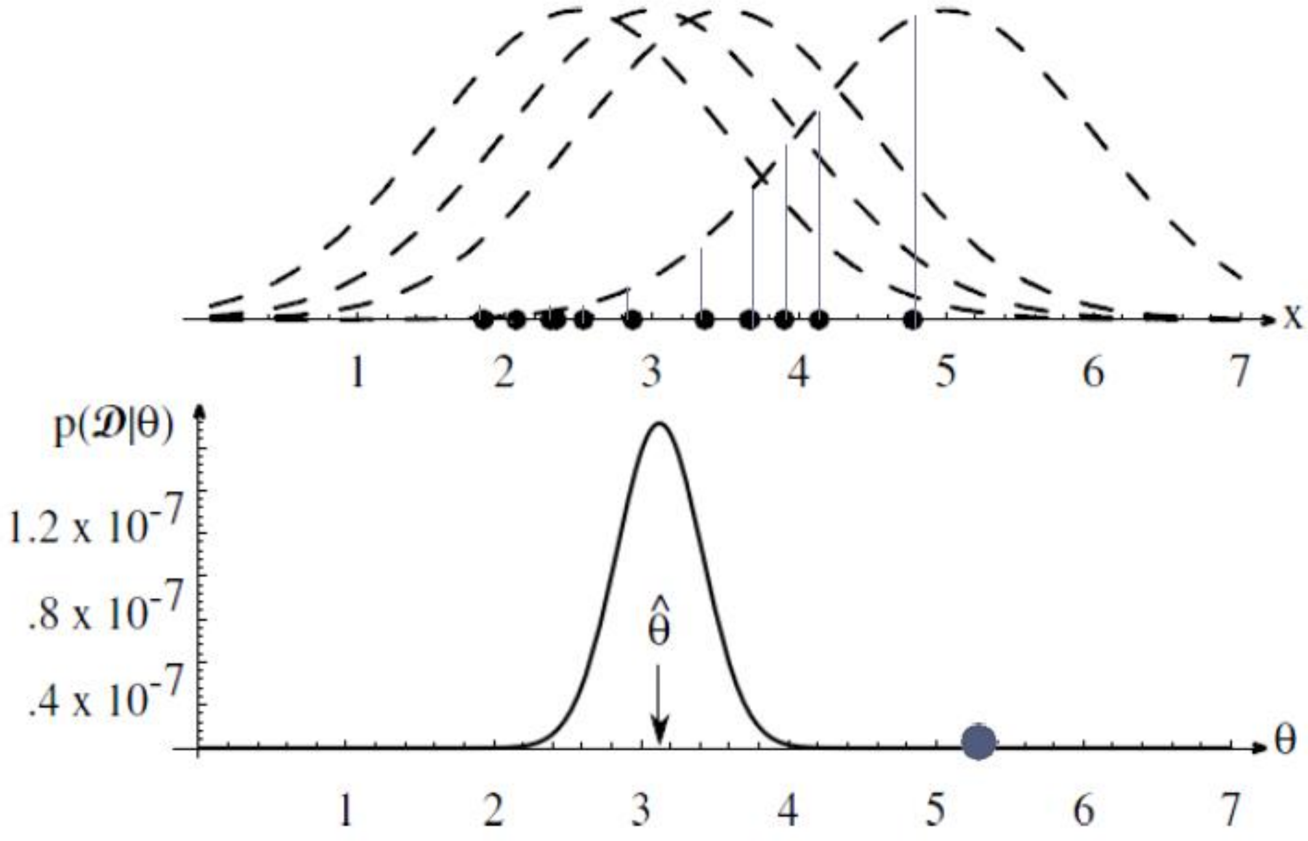
$$P(x|\mu) = N(x|\mu, 1)$$

Maximum Likelihood Estimation (MLE)

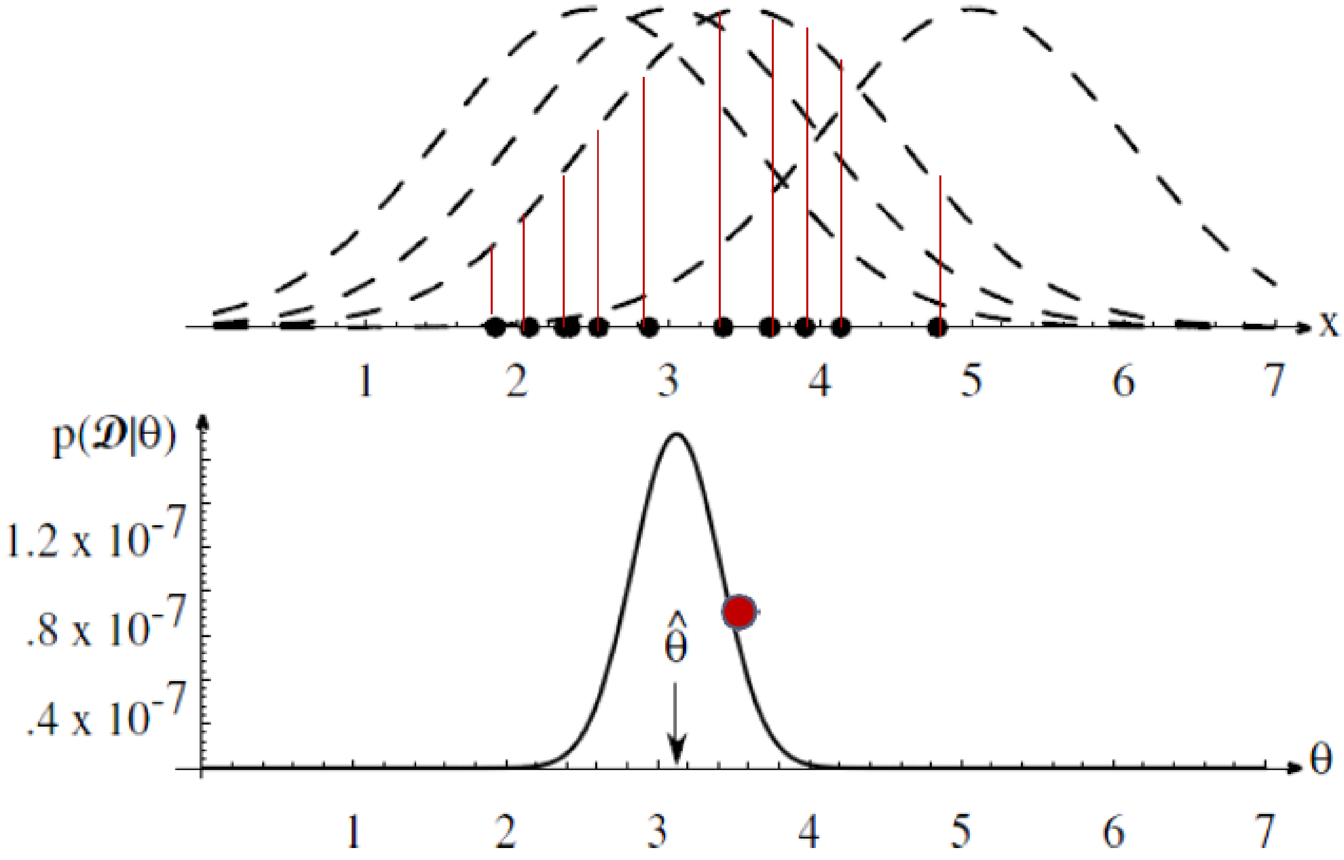
$$P(X|\theta) = \sum_{i=1}^{n} P\left(x^{(i)}\middle|\theta\right)$$

$$\theta_{ML} = argmax\, P(X|\theta)$$

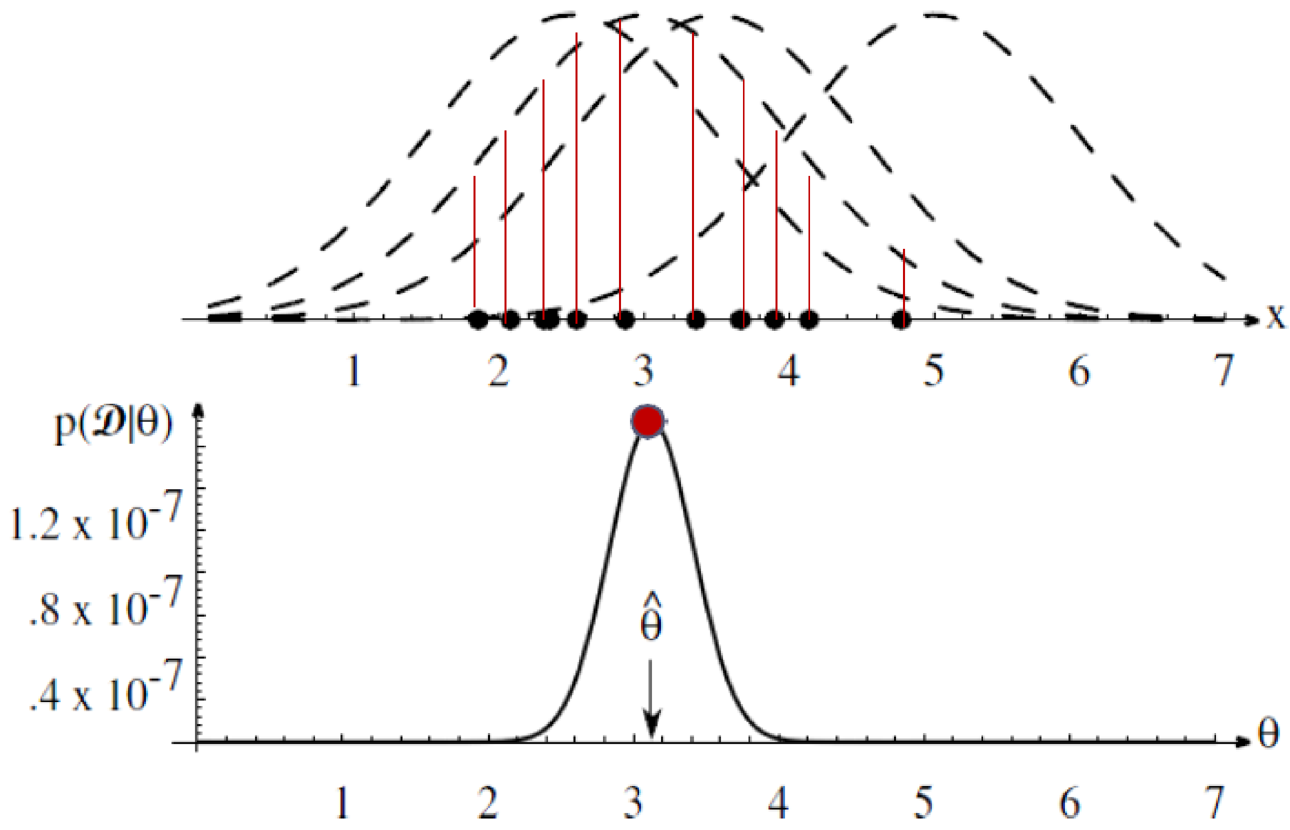
Maximum Likelihood Estimation (MLE)



Maximum A Posteriori (MAP)



Maximum A Posteriori (MAP)



Maximum Likelihood Estimation (MLE)

$$\mathcal{L}(\boldsymbol{\theta}) = \ln p(\mathcal{D}|\boldsymbol{\theta}) = \ln \prod_{i=1}^{N} p\left(\boldsymbol{x}^{(i)}\middle|\boldsymbol{\theta}\right) = \sum_{i=1}^{N} \ln p\left(\boldsymbol{x}^{(i)}\middle|\boldsymbol{\theta}\right)$$

$$\widehat{\boldsymbol{\theta}}_{ML} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathcal{L}(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^{N} \ln p\left(\boldsymbol{x}^{(i)}\middle|\boldsymbol{\theta}\right)$$

MLE Bernoulli

Given: $\mathcal{D} = \{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$, m heads (1), $N - m$ tails (0)

$$p(x|\theta) = \theta^{x}(1-\theta)^{1-x}$$

$$p(\mathcal{D}|\theta) = \prod_{i=1}^{N} p(x^{(i)}|\theta) = \prod_{i=1}^{N} \theta^{x^{(i)}}(1-\theta)^{1-x^{(i)}}$$

$$\ln p(\mathcal{D}|\theta) = \sum_{i=1}^{N} \ln p(x^{(i)}|\theta) = \sum_{i=1}^{N} \{x^{(i)} \ln\theta + (1-x^{(i)}) \ln(1-\theta)\}$$

$$\frac{\partial \ln p(\mathcal{D}|\theta)}{\partial\theta} = 0 \Rightarrow \theta_{ML} = \frac{\sum_{i=1}^{N} x^{(i)}}{N} = \frac{m}{N}$$

Maximum A Posteriori (MAP)

MAP estimation

$$\hat{\boldsymbol{\theta}}_{MAP} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\boldsymbol{\theta}|\mathcal{D})$$

Since $p(\boldsymbol{\theta}|\mathcal{D}) \propto p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})$

$$\hat{\boldsymbol{\theta}}_{MAP} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

Gaussian MAP

$p(x|\mu) \sim N(\mu, \sigma^2)$ μ is the only unknown paramete

$p(\mu|\mu_0) \sim N(\mu_0, \sigma_0^2)$ μ_0 and σ_0 are known

$$\frac{d}{d\mu} \ln \left(p(\mu) \prod_{i=1}^{N} p(x^{(i)}|\mu) \right) = 0$$

$$\Rightarrow \sum_{i=1}^{N} \frac{1}{\sigma^2} (x^{(i)} - \mu) - \frac{1}{\sigma_0^2} (\mu - \mu_0) = 0$$

$$\Rightarrow \hat{\mu}_{MAP} = \frac{\mu_0 + \frac{\sigma_0^2}{\sigma^2} \sum_{i=1}^{N} x^{(i)}}{1 + \frac{\sigma_0^2}{\sigma^2} N}$$

منابع

- Stochastic Processes - Hamid R. Rabiee
- machine learning – Ali Sharifi-Zarchi
- machine learning - Mahdieh Soleymani Baghshah