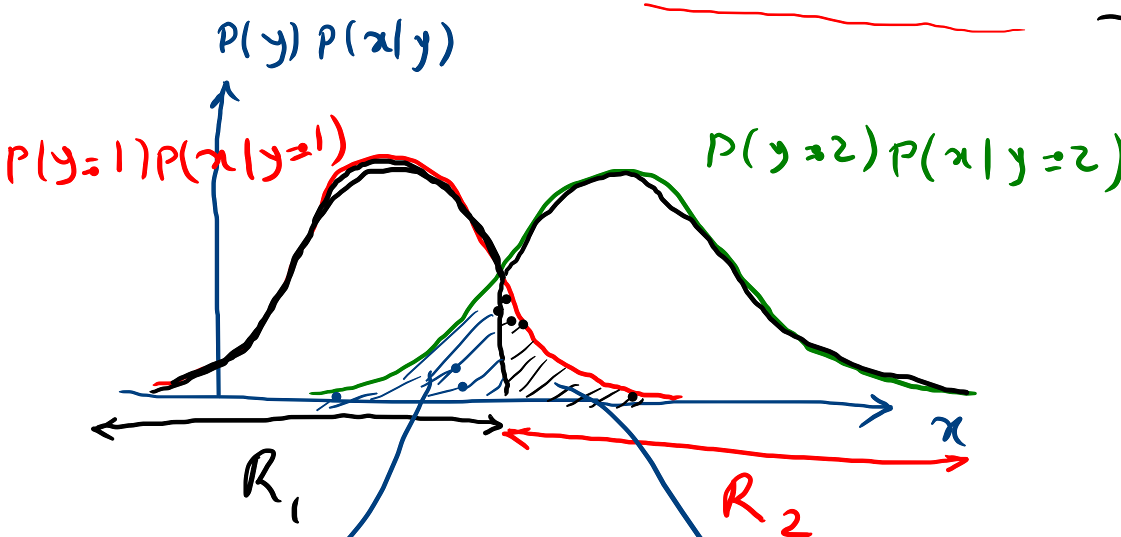
$$P(\text{error}) = P(\hat{y} \neq y) = \underline{P(y=1, x \in R_2)} + \underline{P(y=2, x \in R_1)}$$



$$\int_{R_1} P(y=2)\,P(x|y=2)\,dx$$

$$\int_{R_2} P(y=1)\,P(x|y=1)\,dx$$

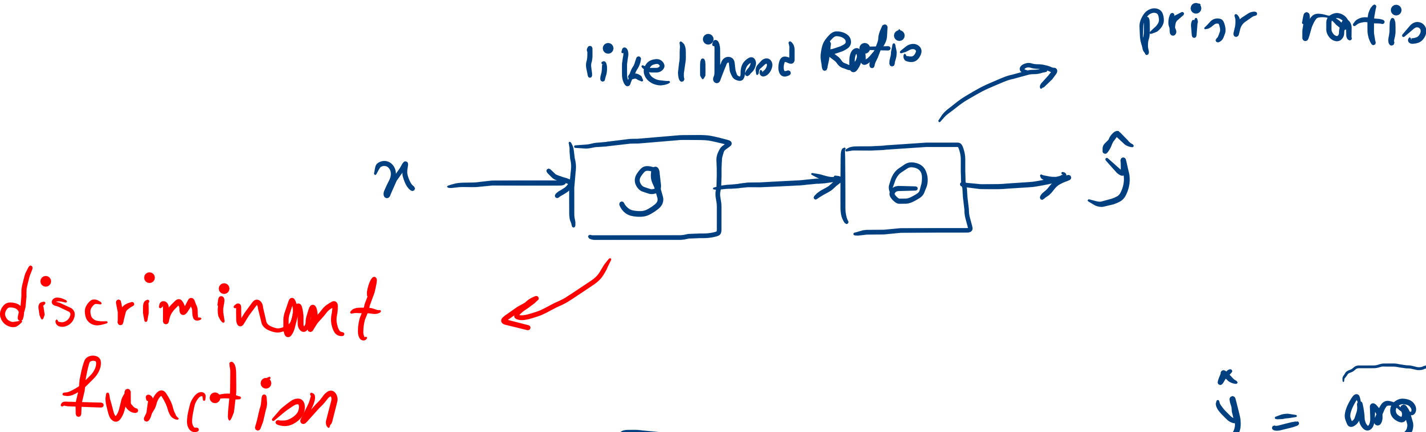
$$P\left(\begin{bmatrix} ۱۷۰ \\ ۷۰ \end{bmatrix} \middle| \; y = \text{مرد} \right)$$

$$P(x \mid y)$$

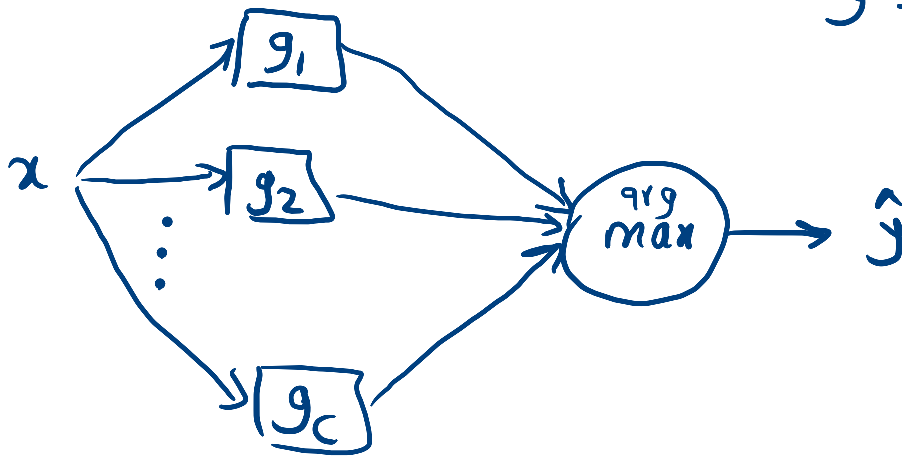
$$P(y \mid x)$$

$$\checkmark \; P\left(y = \text{مرد} \;\middle|\; \begin{bmatrix} ۱۷۰ \\ ۷۰ \end{bmatrix} \right)$$

$$\frac{P(x \mid y = \text{مرد})}{P(x \mid y = \text{زن})} \underset{\text{زن}}{\overset{\text{مرد}}{\gtrless}} \frac{P(y = \text{زن})}{P(y = \text{مرد})} = 1$$



$$\hat{y} = \arg\max_i \underbrace{P(y=i/x)}_{g_i}$$



Actual $\quad\quad$ predicted

$$1 \xrightarrow{\lambda_{21}} 2$$

$$2 \xrightarrow{\lambda_{12}} 1$$

λ_{ij} $\quad$ $i, j \in \{1, \dots, c\}$

predict $\quad$ actual

λ_{11}

$$R(i|x) = \sum_{j=1}^{c} \lambda_{ij} P(y=j|x) = E_{P(y|x)}\left[\lambda(i|y)\right]$$

$$g_i(x) = -R(i|x)$$

$$\boxed{\hat{y} = \arg\min_i R(i|x)}$$

Minimum Risk classifier

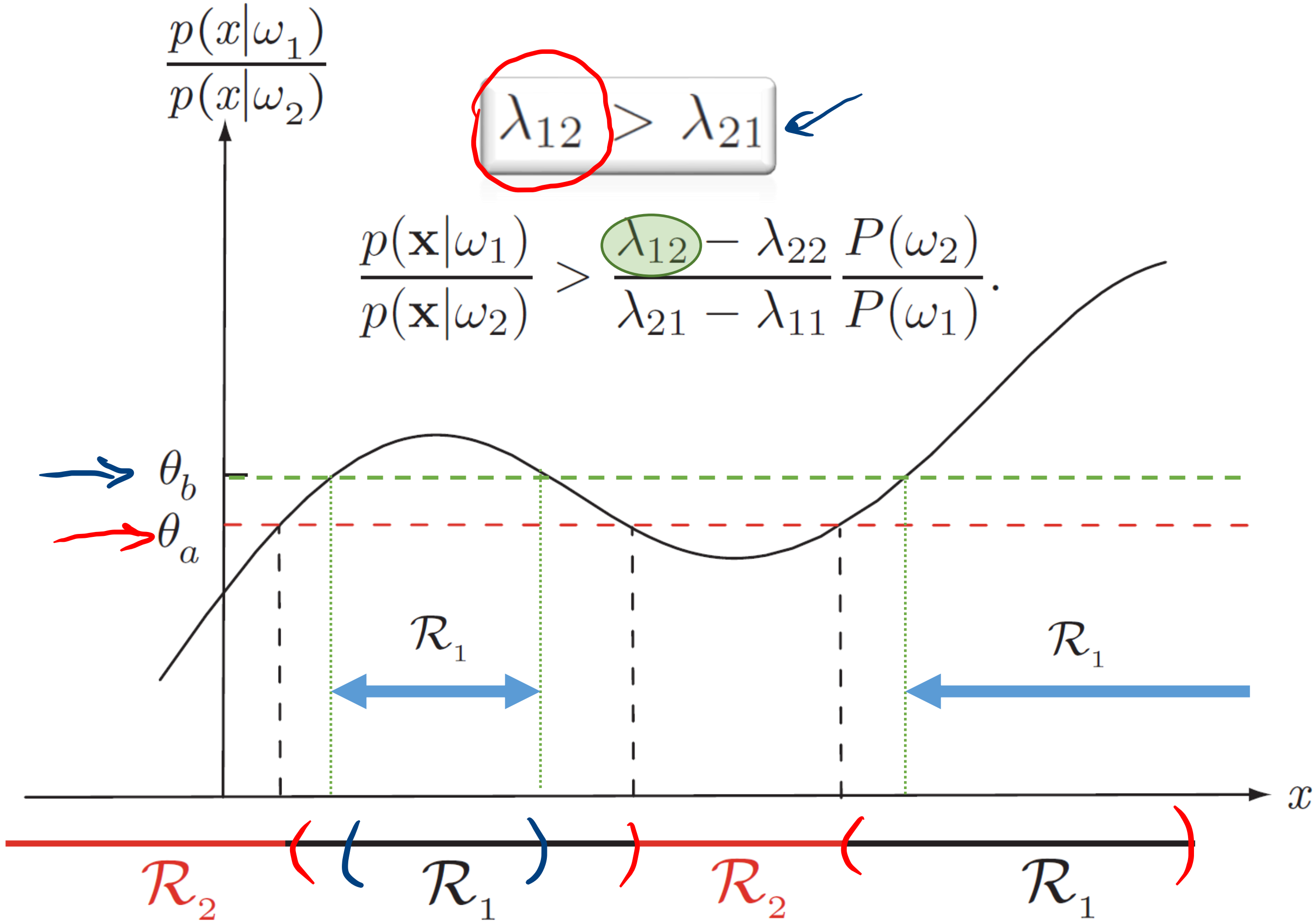
Binary classifier:

$$\frac{R(1|x)}{R(2|x)} \underset{1}{\overset{2}{\gtrless}} 1 \Rightarrow \frac{\lambda_{11} \overbrace{P(y=1|x)}^{\frac{P(x|y=1)P(y=1)}{P(x)}} + \lambda_{12} P(y=2|x)}{\lambda_{21} P(y=1|x) + \lambda_{22} P(y=2|x)} \underset{1}{\overset{2}{\gtrless}} 1$$

$$\underbrace{\overbrace{\frac{P(x|y=1)}{P(x|y=2)}}^{g(x)}}_{\text{likelihood Ratio}} \underset{1}{\overset{2}{\gtrless}} \overbrace{\underbrace{\frac{\lambda_{22}-\lambda_{12}}{\lambda_{11}-\lambda_{21}}}\ \underbrace{\frac{P(y=2)}{P(y=1)}}_{\text{prior ratio}}}^{\theta}$$

$$g(x) \underset{2}{\overset{1}{\gtrless}} \theta$$

Minimum-Error-Rate Classification



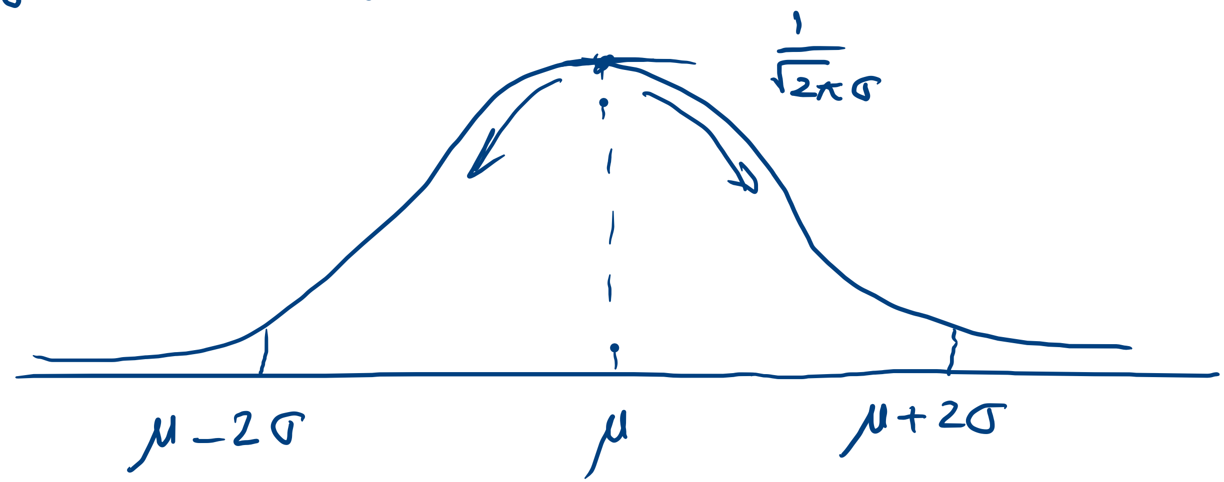
The likelihood ratio $p(\mathbf{x}|w_1)/p(\mathbf{x}|w_2)$. The threshold θ_a is computed using the priors $P(w_1) = 2/3$ and $P(w_2) = 1/3$, and a zero-one loss function. If we penalize mistakes in classifying w_2 patterns as w_1 more than the converse, we should increase the threshold to θ_b.

Normal Distribution: $\quad x \in \mathbb{R}$

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

$$\mu = E[x]$$

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



Multi-variate Normal Distribution: $x \in \mathbb{R}^d$

$$\underline{\underline{x}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

$$P(x) = \frac{1}{(2\pi)^{d/2} \underline{|\Sigma|}^{1/2}} \exp\left\{ -\frac{1}{2} \underbrace{\underset{1\times d}{(x-\mu)^T} \underset{d\times d}{\Sigma^{-1}} \underset{d\times 1}{(x-\mu)}}_{\text{scalar}} \right\}$$

$$= \mathcal{N}\left(x \,\middle|\, \underset{d\times 1}{\mu}, \underset{d\times d}{\Sigma}\right)$$

$$\mu = E[x] = \begin{bmatrix} E[x_1] \\ \vdots \\ E[x_d] \end{bmatrix}$$

$$\Sigma = E\left[(x-\mu)(x-\mu)^T\right]$$

$$\Sigma_{ij} = E\left[(x_i-\mu_i)(x_j-\mu_j)\right]$$

$$x^T A x = \sum_i \sum_j x_i x_j A_{ij}$$

$$x^T A^T x = \sum_i \sum_j x_i x_j A_{ji}$$

$$x^T A x = x^T \overbrace{\underbrace{\left(\frac{A + A^T}{2} \right)}_{\text{متقارن}}}^{B} x = x^T B x$$

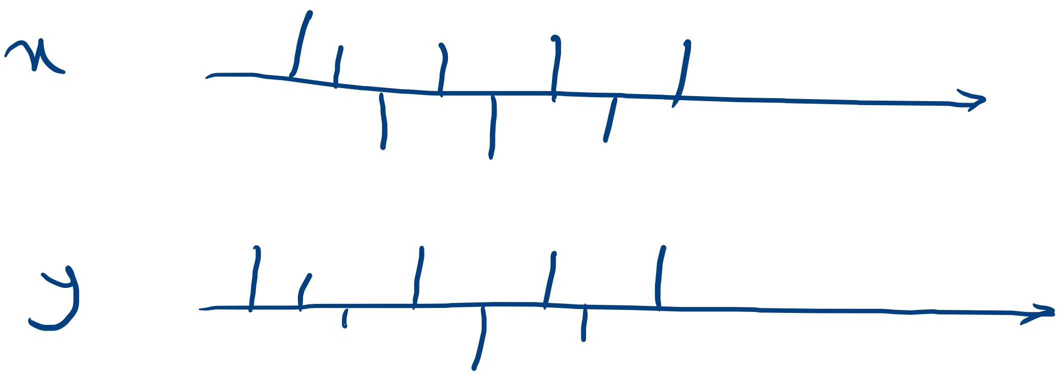
$$\Sigma_{ij} = E\left[(x_i - \mu_i)(x_j - \mu_j)\right]$$

$$E[xy] \Longrightarrow \text{correlation}$$

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

$$\text{correlation} \overset{\times}{\Longrightarrow} \text{indp.}$$

$$\Longleftarrow \checkmark$$

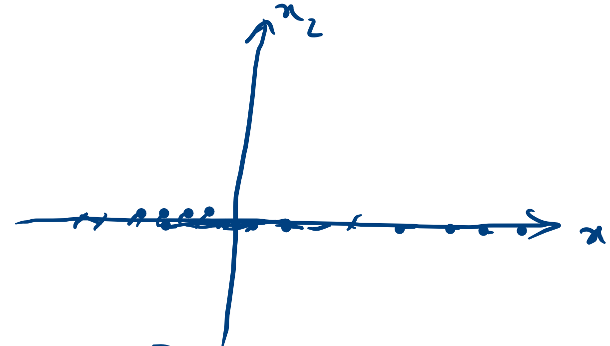


$$E[xy] = \iint xy\, \underbrace{P(x,y)}\, dx\, dy$$

$$\underbrace{\int x\, P(x)\, dx}_{E[X]} \; \underbrace{\int y\, P(y)\, dy}_{E[Y]}$$

$$= 0$$

$$x^T \underline{\underline{A}}\, \underline{\underline{x}} = \sum_i \sum_j x_i x_j \underline{\underline{A_{ij}}}$$



$$\underline{\underline{x}}^T \underline{\underline{x}} = \sum_{i=1}^{d} x_i^2$$

$$\cancel{\exp\left\{-\frac{1}{2}(x-\mu)\frac{1}{\sigma}(x-\mu)\right\}}$$

$\underline{\underline{\Sigma}}$ positive definite

$$A^T = A$$

$$x^T \Sigma x \geq 0$$

$$x = bb$$

$$\boxed{A = B^T B} \Rightarrow x^T A x > 0$$

$$x^T A x = x^T B^T B x = (Bx)^T (Bx) = \|Bx\|_2^2 > 0$$

$$x \neq 0$$