Linear Discrimination

Classification $\Rightarrow$ $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$ $\quad y_i \in \{1, 2, \dots, K\}$

$$\left.\begin{array}{l} g_1(x) \\ g_2(x) \\ \vdots \\ g_K(x) \end{array}\right\} \text{score functions}$$

choose c^* if $g_{c^*}(x) = \max_{c=1}^{K} g_c(x)$

$$g_c(x) = p(x \mid y=c) \cdot \boxed{P(y=c)}$$

Univariate $(x_i \in \mathbb{R})$

μ_c, σ_c^2

$\Downarrow$

$\hat{\mu}_c, \hat{\sigma}_c^2$

multivariate $(x_i \in \mathbb{R}^D)$

μ_c, Σ_c

$\Downarrow$

$\hat{\mu}_c, \hat{\Sigma}_c$

$D \times 1 \leftarrow \hat{\mu}_c$, $\quad \hat{\Sigma}_c \rightarrow D \times D$

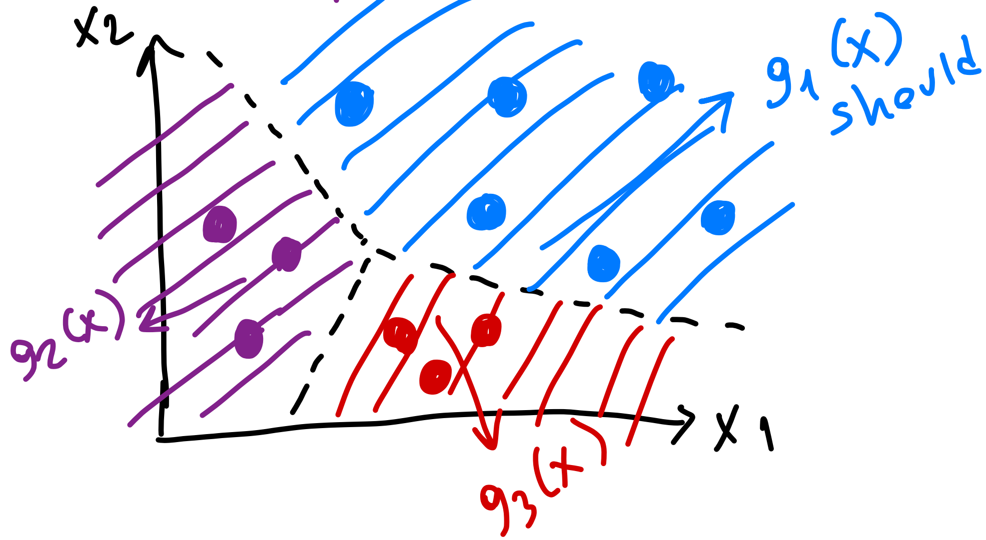
$$\frac{N_c}{N} = \frac{\text{\# of data points from class } c}{\text{total \# data point}}$$

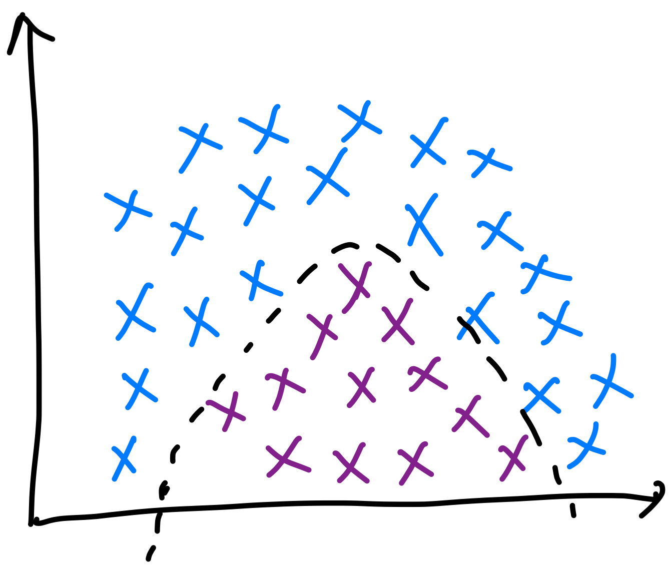
$$g_c(x \mid w_c, w_{c0}) = \underbrace{w_c^T}_{1 \times D} \cdot \underbrace{x}_{D \times 1} + \underbrace{w_{c0}}_{1 \times 1} = \sum_{d=1}^{D} w_{cd} \cdot x_d + w_{c0}$$

$$\begin{bmatrix} w_{c1} & w_{c2} & \cdots & w_{cD} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} + w_{c0} = w_{c1} \cdot x_1 + w_{c2} \cdot x_2 + \cdots + w_{c0}$$

$$\boxed{\begin{array}{l} 2 \cdot x \cdot y + 4 \cdot y \cdot x \\ 0 \cdot x \cdot y + 6 \cdot y \cdot x \\ 3\, x \cdot y + 3\, y \cdot x \end{array}}$$

total # of parameters : $K(D+1)$





$$g_c(x \mid \underbrace{W_c}_{D \times D}, \underbrace{w_c}_{D \times 1}, \underbrace{w_{c0}}_{1 \times 1}) = x^T \cdot W_c \cdot x + w_c^T \cdot x + w_{c0}$$

total # of parameters: $K\left(\frac{D \cdot (D+1)}{2} + D + 1\right)$

Binary Classification $(K=2)$

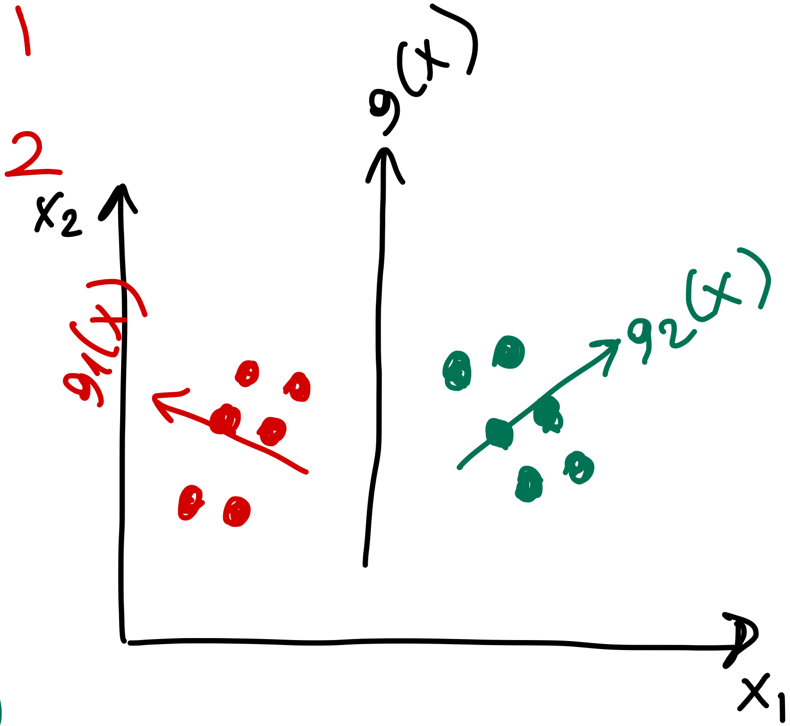
$$\left.\begin{matrix} g_1(x) \\ g_2(x) \end{matrix}\right\} \begin{matrix} \text{if } \underline{g_1(x) > g_2(x)} \Rightarrow \hat{y} = 1 \\ \text{if } \underline{g_2(x) > g_1(x)} \Rightarrow \hat{y} = 2 \end{matrix}$$

$$\text{if } \underline{g_1(x) - g_2(x)} > 0 \Rightarrow \hat{y} = 1$$

$$\text{if } \underline{g_1(x) - g_2(x)} < 0 \Rightarrow \hat{y} = 2$$

$$\text{if } g(x) > 0 \Rightarrow \hat{y} = 1$$

$$\text{if } g(x) < 0 \Rightarrow \hat{y} = 2$$



$$\begin{aligned} g_1(x) &= w_1^T . x + w_{10} \\ - \quad g_2(x) &= w_2^T . x + w_{20} \\ \hline g_1(x) - g_2(x) &= w_1^T . x - w_2^T . x + w_{10} - w_{20} \\ &= \underbrace{(w_1 - w_2)}^T . x + (\underbrace{w_{10} - w_{20}}) \\ &= w^T . x + w_0 \end{aligned}$$

$$g_1(x) = x^T \cdot \underline{W_1} \cdot x + w_1^T \cdot x + w_{10}$$

$$x^T \cdot A x - x^T \cdot B x = x^T \cdot (A - B) \cdot x$$

$$- \quad g_2(x) = x^T \cdot \underline{W_2} \cdot x + w_2^T \cdot x + w_{20}$$

$$g_1(x) - g_2(x) = x^T \cdot W_1 x - x^T \cdot W_2 \cdot x + w_1^T \cdot x - w_2^T \cdot x + w_{10} - w_{20}$$

$$= x^T \cdot \underbrace{(W_1 - W_2)}_{W} \cdot x + \underbrace{(w_1 - w_2)}_{w}^T \cdot x + \underbrace{(w_{10} - w_{20})}_{w_0}$$

$$g(x) = x^T \cdot W \cdot x + w^T \cdot x + w_0$$

Multiclass Classification ($K > 2$)

$$\left.\begin{matrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_K(x) \end{matrix}\right\} K = 3$$

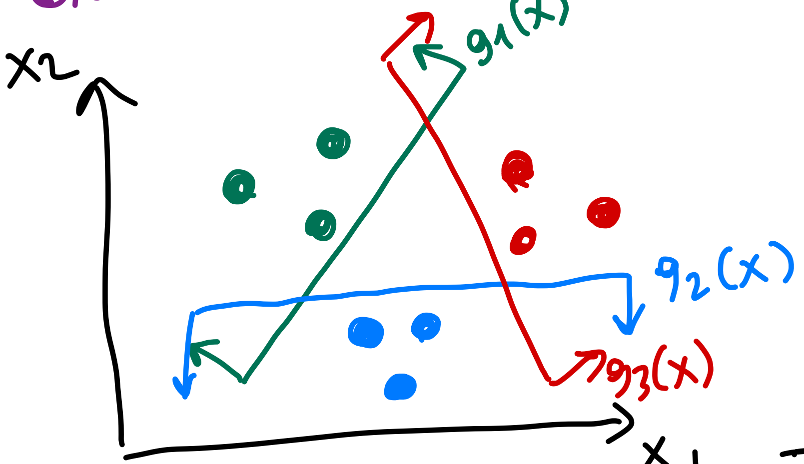
if $g_1(x) > g_2(x)$ & $g_1(x) > g_3(x) \Rightarrow \hat{y} = 1$

if $g_2(x) > g_1(x)$ & $g_2(x) > g_3(x) \Rightarrow \hat{y} = 2$

if $g_3(x) > g_1(x)$ & $g_3(x) > g_2(x) \Rightarrow \hat{y} = 3$

$$\hat{y} = \arg\max_{c=1}^{K} g_c(x)$$

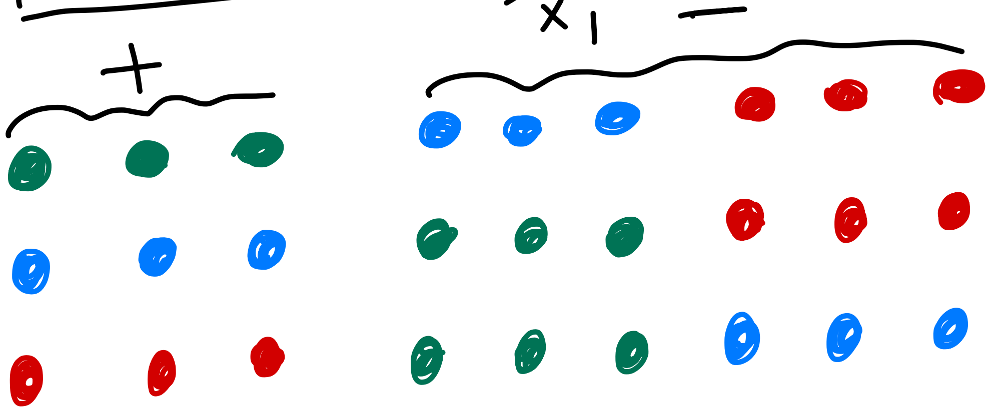
ONE-VERSUS-ALL (OVA) APPROACH



3-class problem

$\Downarrow$

3 binary-classification problem



$g_1(x)$ Green vs nongreen

$g_2(x)$ Blue vs non blue

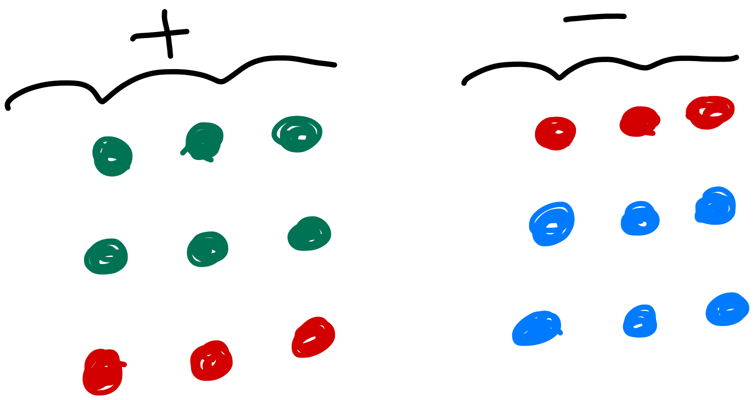
$g_3(x)$ Red vs nonred

K classes $\Rightarrow$ # of parameters $= K(D+1)$

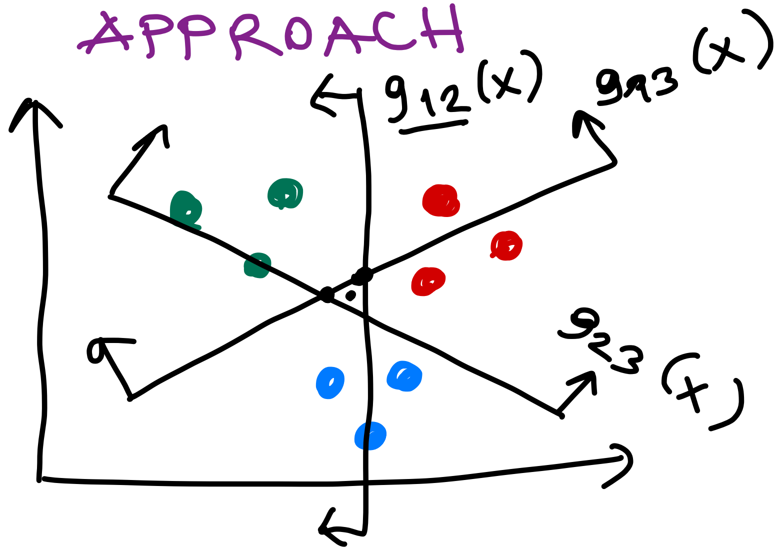
data set size for each classification problem $= N$

# of score functions $= K$

ONE-VERSUS-OTHER (OVO) APPROACH



$g_{12}(x)$
$g_{13}(x)$
$g_{23}(x)$



x^*	1	2	3
$g_{12}(x^*)$	1	0	X
$g_{13}(x^*)$	1	X	0
$g_{23}(x^*)$	X	0	1
# of wins	2	0	1

$\boxed{1 \quad 1 \quad 1}$

"ties are broken arbitrarily"

K classes

$$\# \text{ of parameters} = \frac{K(K-1)}{2}(D+1)$$

$$\text{data set size for each problem} = \frac{2N}{K}$$

$$\# \text{ of score functions} = \frac{K(K-1)}{2}$$

$K=2$

$$P(y=1 \mid x) = \delta$$
$$P(y=2 \mid x) = 1-\delta$$

$$\begin{cases} \text{if } \delta > 0.5 \Rightarrow \hat{y} = 1 \\ \text{if } \dfrac{\delta}{1-\delta} > 1 \Rightarrow \hat{y} = 1 \\ \text{if } \log\left(\dfrac{\delta}{1-\delta}\right) > 0 \Rightarrow \hat{y} = 1 \end{cases}$$

$$\log\left[\frac{P(y=1 \mid x)}{P(y=2 \mid x)}\right] = \log\left[\frac{\dfrac{P(x \mid y=1)\,P(y=1)}{\cancel{P(x)}}}{\dfrac{P(x \mid y=2)\,P(y=2)}{\cancel{P(x)}}}\right]$$

$$\log\left[\frac{a \cdot b}{c \cdot d}\right] = \log\left[\frac{a}{c}\right] + \log\left[\frac{b}{d}\right]$$

$$= \log\left[\frac{\overbrace{P(x \mid y=1)}^{N(x;\,\mu_1,\,\Sigma)}}{\underbrace{P(x \mid y=2)}_{N(x;\,\mu_2,\,\Sigma)}}\right] + \underline{\log\left[\frac{P(y=1)}{P(y=2)}\right]}$$

$$p(x \mid y=1) = N(x;\ \underbrace{\mu_1}_{D\times 1},\ \underbrace{\Sigma_1}_{D\times D})$$
$$p(\underbrace{x}_{\mathbb{R}^D} \mid y=2) = N(x;\ \underbrace{\mu_2}_{D\times 1},\ \underbrace{\Sigma_2}_{D\times D})$$

$$\Sigma_1 = \Sigma_2 = \Sigma \quad (\text{equal covariance assumption})$$

$$\frac{1}{\sqrt{(2\pi)^D \cdot |\Sigma|}} \cdot \exp\left[-\frac{1}{2}(x-\mu)^T \cdot \Sigma^{-1} \cdot (x-\mu)\right] = \mathcal{N}(x;\mu,\Sigma)$$

$$\frac{\exp(a)}{\exp(b)} = \exp(a-b)$$

$$\log\left[\frac{\cancel{(2\pi)^{-D/2}} \cdot \cancel{|\Sigma|^{-1/2}}\, \exp\left[-\frac{1}{2}(x-\mu_1)^T \cdot \Sigma^{-1}(x-\mu_1)\right]}{\cancel{(2\pi)^{-D/2}} \cdot \cancel{|\Sigma|^{-1/2}}\, \exp\left[-\frac{1}{2}(x-\mu_2)^T \cdot \Sigma^{-1}(x-\mu_2)\right]}\right] + \log\left[\frac{P(y=1)}{P(y=2)}\right]$$

$$= -\frac{1}{2}(x-\mu_1)^T \cdot \Sigma^{-1} \cdot (x-\mu_1) + \frac{1}{2}(x-\mu_2)^T \cdot \Sigma^{-1}(x-\mu_2) + \log\left[\frac{P(y=1)}{P(y=2)}\right]$$

$$= \cancel{-\frac{1}{2}x^T \cdot \Sigma^{-1} \cdot x} + \mu_1^T \cdot \Sigma^{-1} \cdot x - \frac{1}{2}\mu_1^T \cdot \Sigma^{-1} \cdot \mu_1$$
$$\cancel{+\frac{1}{2}x^T \cdot \Sigma^{-1} \cdot x} - \mu_2^T \Sigma^{-1} \cdot x + \frac{1}{2}\mu_2^T \cdot \Sigma^{-1} \cdot \mu_2 + \log\left[\frac{P(y=1)}{P(y=2)}\right]$$

Exercise #3

$$= \underbrace{(\mu_1-\mu_2)^T \cdot \Sigma^{-1}}_{w^T} \cdot x + \underbrace{\left[-\frac{1}{2}(\mu_1+\mu_2)^T \cdot \Sigma^{-1}(\mu_1-\mu_2) + \log\left[\frac{P(y=1)}{P(y=2)}\right]\right]}_{w_0}$$

$$= w^T \cdot x + w_0 \quad \Big\}\ \text{where} \quad w = \Sigma^{-1} \cdot (\mu_1 - \mu_2)$$

$\Sigma \Rightarrow \hat{\Sigma}$, $\mu_1 \Rightarrow \hat{\mu}_1$ $\quad \mu_2 \Rightarrow \hat{\mu}_2$ $\quad \hat{P}(y=1)$ $\quad \hat{P}(y=2)$