Linear Discrimination

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

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

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

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

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

μ_c, σ_c^2

$\Downarrow$

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

$\hat{\mu}_c$: 1×1, $\hat{\sigma}_c^2$: 1×1

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

μ_c, Σ_c

$\Downarrow$

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

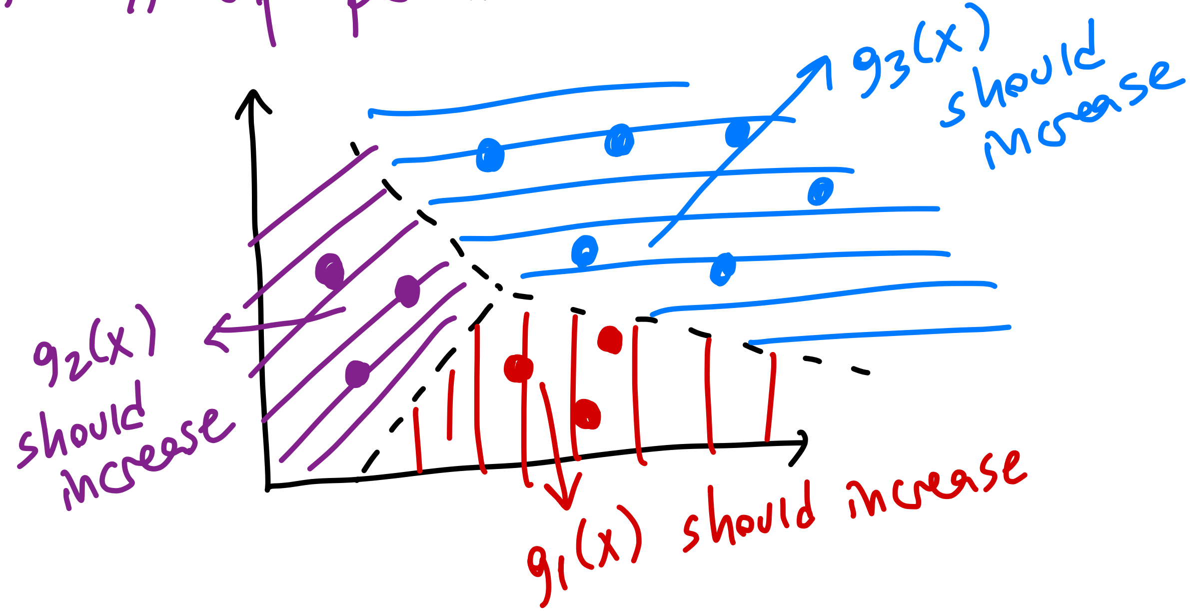
$\hat{\mu}_c$: $D \times 1$, $\hat{\Sigma}_c$: $D \times D$

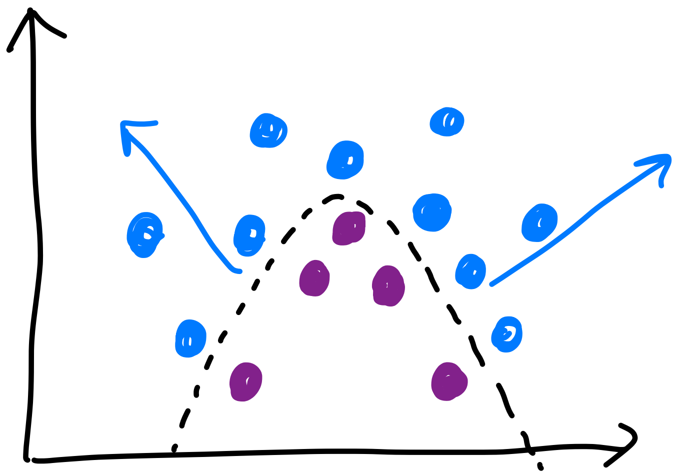
$\dfrac{N_c}{N} = \dfrac{\text{\# of data points in class \#c}}{\text{total \# of data points}}$

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

$$= \sum_{d=1}^{D}[w_{cd}\,x_d] + w_{c0} = [w_{c1}\ w_{c2}\ \cdots\ w_{cD}]\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_D\end{bmatrix} + w_{c0}$$

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





$$g_c(x \mid W_c, w_c, w_{c0}) = x^T W_c x + w_c^T \cdot x + w_{c0}$$

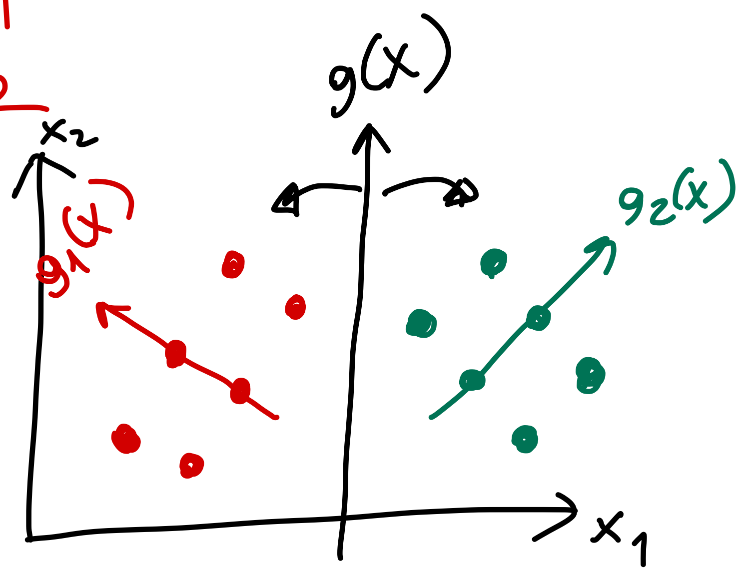
total # of parameters = $K\left[\frac{D \cdot (D+1)}{2} + D + 1\right]$

Binary classification $(K=2)$

$$\left.\begin{array}{l} g_1(x) \\ g_2(x) \end{array}\right\} \begin{array}{lll} \text{if } g_1(x) > g_2(x) & \Rightarrow & \hat{y}=1 \\ \text{if } g_2(x) > g_1(x) & \Rightarrow & \hat{y}=2 \end{array}$$

$$\text{if } \underline{g_1(x)-g_2(x)} > 0 \Rightarrow \hat{y}=1$$
$$\text{if } \underline{g_1(x)-g_2(x)} < 0 \Rightarrow \hat{y}=2$$

$$\text{if } g(x) > 0 \Rightarrow \hat{y}=1$$
$$\text{if } g(x) < 0 \Rightarrow \hat{y}=2$$



$$g_1(x) = w_1^T . x + w_{10}$$
$$- \quad g_2(x) = w_2^T . x + w_{20}$$

$$g_1(x)-g_2(x) = w_1^T . x - w_2^T . x + w_{10} - w_{20}$$
$$= (w_1 - w_2)^T . x + (w_{10} - w_{20}) = w^T . x + w_0$$

$$g_1(x) = x^T . W_1 x + w_1^T . x + w_{10}$$

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

$$x^T . A . x - x^T . B . x = x^T (A - B) . x$$

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

$$= x^T . (W_1 - W_2) . x + (w_1 - w_2)^T . x + (w_{10} - w_{20})$$

$$\boxed{g(x) = x^T . \underbrace{W}_{D \times D} x + \underbrace{w}_{1 \times D}^T . x + \underbrace{w_0}_{1 \times 1}}$$

Multiclass Classification $(K > 2)$

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

assume $K = 3$

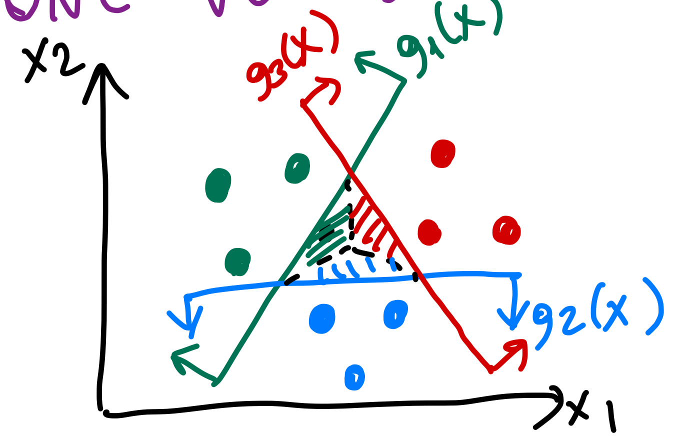
if $g_1(x) > g_2(x)$ & $g_1(x) > g_3(x) \Rightarrow \hat{y} = 1$

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

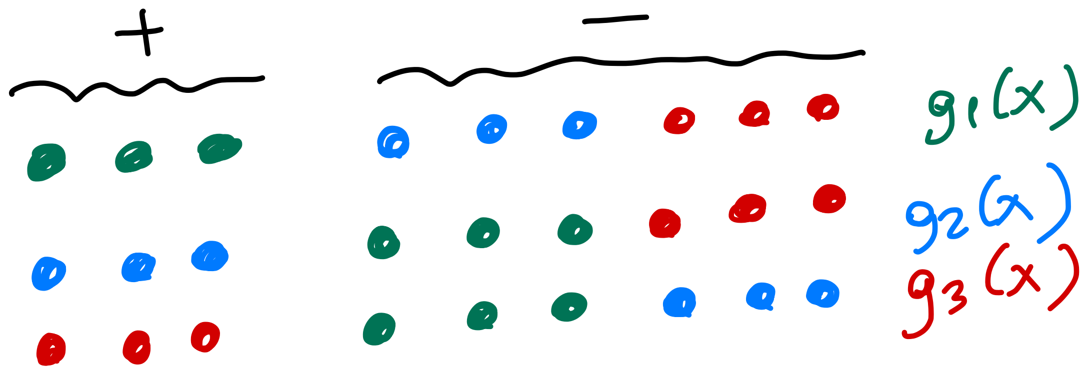
if $g_3(x) > g_1(x)$ & $g_3(x) > g_2(x) \Rightarrow \hat{y} = 3$

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

ONE-VERSUS-ALL (OVA) APPROACH



3-class problem $\Rightarrow$ 3 binary classification problems

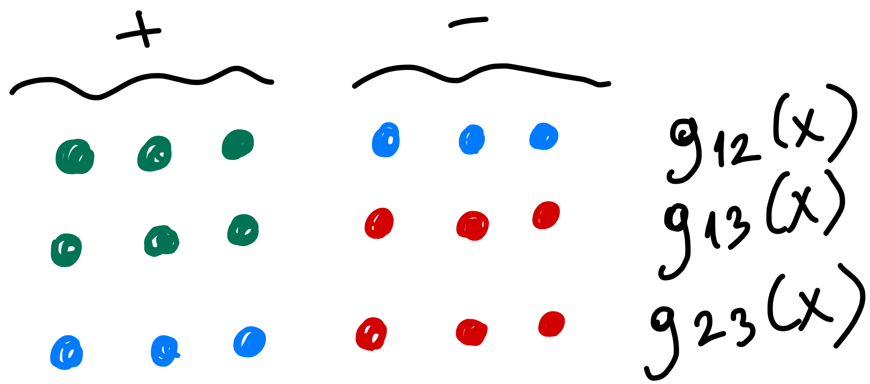


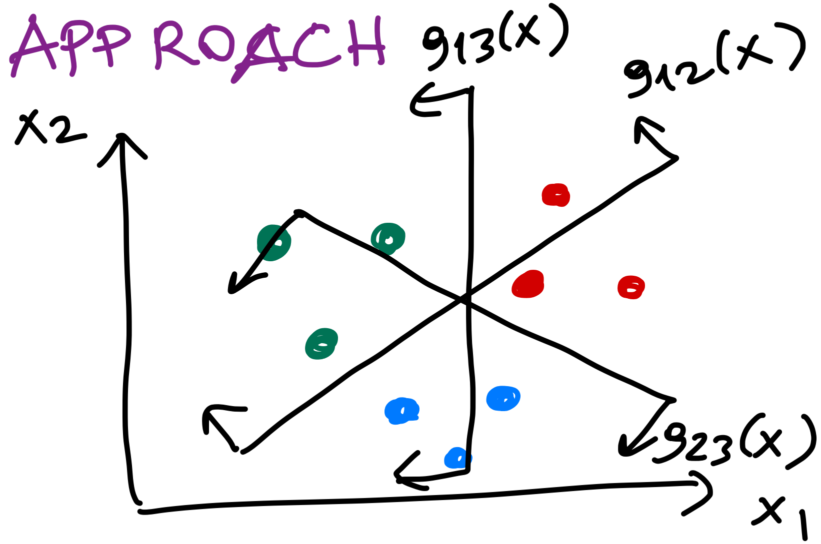
K classes $\Rightarrow$ # of parameters $= K(D+1)$

data set size for each problem $= N$

# of score functions $= K$

ONE-VERSUS-OTHER (OVO) APPROACH





x	1	2	3
$g_{12}(x)$	1	0	$\emptyset$
$g_{13}(x)$	1	$\emptyset$	0
$g_{23}(x)$	$\emptyset$	0	1
# of wins	2	0	1

$\boxed{\hat{y} = 1}$

K classes

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

data set size for each problem $= \frac{2.N}{K}$ → assuming equal frequencies

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

if there is a tie ⇒ "ties are broken arbitrarily" or another tie-breaking rule.

$K=2$

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

$$\begin{cases} \text{if } \delta > 0.5 \Rightarrow \hat{y}=1 \\ \text{if } \frac{\delta}{1-\delta} > 1 \Rightarrow \hat{y}=1 \\ \text{if } \log\left[\frac{\delta}{1-\delta}\right] > 0 \Rightarrow \hat{y}=1 \end{cases}$$

$$\log\left[\frac{P(y=1|x)}{P(y=2|x)}\right] = \log\left[\frac{\frac{p(x|y=1)P(y=1)}{\cancel{p(x)}}}{\frac{p(x|y=2)P(y=2)}{\cancel{p(x)}}}\right] = \log\left[\frac{p(x|y=1)}{p(x|y=2)}\right] + \log\left[\frac{P(y=1)}{P(y=2)}\right]$$

$$p(x|y=1) \to N(x;\mu_1,\Sigma) \qquad p(x|y=2) \to N(x;\mu_2,\Sigma)$$

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

$$p(x|y=1)=N(x;\overbrace{\mu_1}^{D\times 1},\overbrace{\Sigma_1}^{D\times D})$$

$$p(x|y=2)=N(x;\underbrace{\mu_2}_{D\times 1},\underbrace{\Sigma_2}_{D\times D})$$

$\Sigma_1=\Sigma_2=\Sigma$ (equal covariance assumption)

$$\boxed{\exp(a)/\exp(b) = \exp(a-b)}$$

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

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

$$\boxed{(AB)^T = B^T \cdot A^T}$$

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

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

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

$$= W^T \cdot x + w_0$$

$$w = \overbrace{\Sigma^{-1}}^{D \times D} \cdot \overbrace{(\mu_1-\mu_2)}^{D \times 1}$$

$$\boxed{\begin{array}{ll} \mu_1 \Rightarrow \hat{\mu}_1 & P(y=1) \Rightarrow \hat{P}(y=1) \\ \mu_2 \Rightarrow \hat{\mu}_2 & P(y=2) \Rightarrow \hat{P}(y=2) \\ \Sigma \Rightarrow \hat{\Sigma} \text{ (whole data set)} & \end{array}}$$