Linear Discrimination (chapter #10)

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

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

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

$$g_c(x) = \underbrace{\hat{p}(x|y=c)}_{\text{Chapter \#4}} \; \overbrace{\hat{p}(y=c)}^{\frac{\text{\# of points in class } c}{\text{\# of points}}}$$

chapter #5

univariate: $\mu_c, \sigma_c^2 \Downarrow \hat{\mu}_c, \widehat{\sigma_c^2}$

multivariate ($x \in \mathbb{R}^D$): $D\times 1 \leftarrow \mu_c, \Sigma_c \rightarrow D \times D$ $\Downarrow$ $\hat{\mu}_c, \hat{\Sigma}_c$

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

$$\begin{bmatrix} w_{c1} \\ \vdots \\ w_{cd} \\ \vdots \\ w_{cD} \end{bmatrix} \leftarrow d\text{th entry.}$$

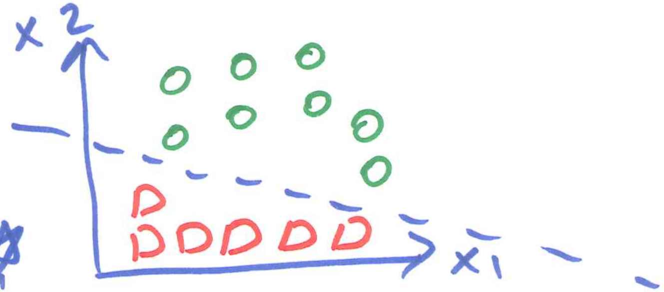
$$= \begin{bmatrix} w_{c1} & w_{c2} & \cdots & w_{cD} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix} + w_{c0}$$

$$= w_{c1} \cdot x_1 + w_{c2} \cdot x_2 + \ldots + w_{cD} \cdot x_D + w_{c0}$$

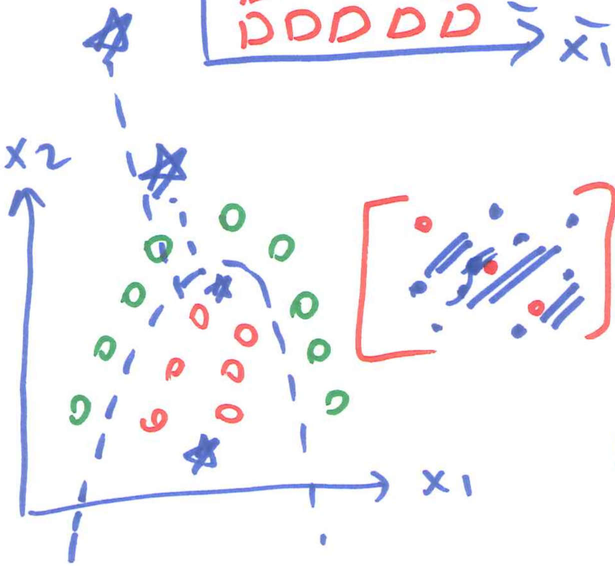
K-classification

\# of parameters

$= (D+1)K$



$$\underbrace{\begin{bmatrix} 1 & 0.6 \\ 0.4 & 1 \end{bmatrix}}_{W_c} \qquad \underbrace{\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}}_{W_c}$$



$$g_c(x \mid W_c, w_c, w_{c0})$$

$$= \underbrace{x^T \cdot W_c \cdot x}_{} + w_c^T \cdot x + w_{c0}$$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} W_{c1,1} & W_{c1,2} \\ W_{c2,1} & W_{c2,2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$x_1 x_2 \cdot 0.6$

$+ x_2 \cdot x_1 \cdot 0.4$

$x_1 \cdot x_2 \cdot 0.5 +$

$x_2 \cdot x_1 \cdot 0.5$

$$W_{c1,1} x_1^2 + W_{c2,2} \cdot x_2^2 + 2 x_1 \cdot x_2 \cdot W_{c1,2} + \cancel{x_2 \cdot x_1 \cdot W_{c2,1}}$$

Two-class classification $(K=2)$

$g_1(x) \quad g_2(x)$

$$\boxed{\begin{array}{l} y=1 \text{ if } g_1(x) > g_2(x) \\ y=2 \text{ if } g_2(x) > g_1(x) \end{array}}$$

$$\boxed{\begin{array}{l} y=1 \text{ if } g_1(x) - g_2(x) > 0 \\ y=2 \text{ if } g_1(x) - g_2(x) < 0 \end{array}}$$

where $g(x) = g_1(x) - g_2(x)$

$$\boxed{\begin{array}{l} y=1 \text{ if } g(x) > 0 \\ y=2 \text{ if } g(x) < 0 \end{array}}$$

$$\left.\begin{array}{l} g_1(x) = w_1^T.x + w_{10} \\ g_2(x) = w_2^T.x + w_{20} \end{array}\right\} \Rightarrow g(x) = w_1^T x + w_{10} - w_2^T x - w_{20}$$

$$g(x) = \underbrace{(w_1 - w_2)}_{w}^T.x + \underbrace{(w_{10} - w_{20})}_{w_0}$$

$$\theta = \underbrace{\{w_1, w_2, w_{10}, w_{20}\}}_{2(D+1)}$$

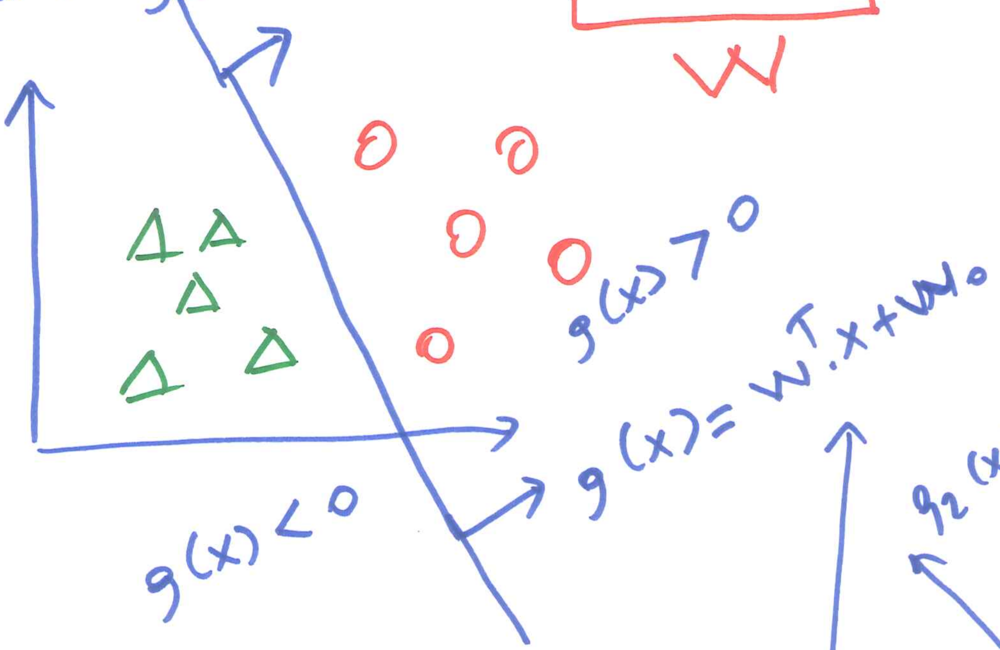
$$\theta = \underbrace{\{w, w_0\}}_{D+1}$$

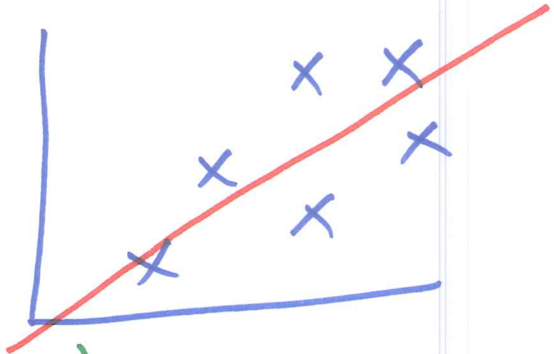
$$\boxed{g(x) = w^T.x + w_0}$$

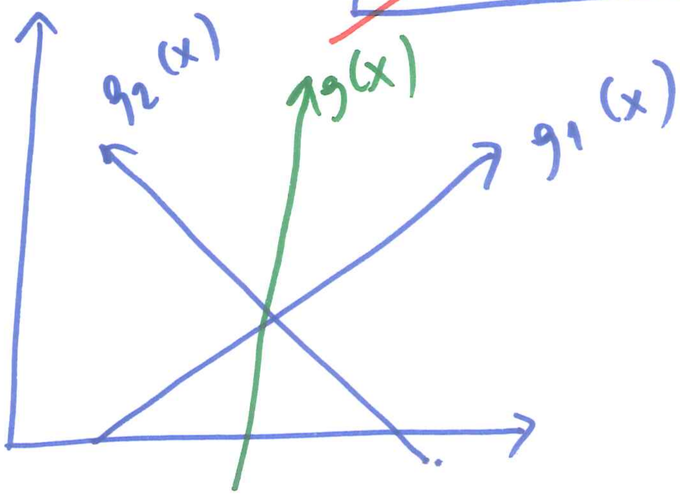
$$g_1(x) = x^T . W_1 x + w_1^T x + w_{10}$$

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

$$g_1(x) - g_2(x) = x^T \underbrace{(W_1 - W_2)}_{W} x + (\underbrace{w_1 - w_2}_{w})^T x + (\underbrace{w_{10} - w_{20}}_{w_0})$$







MULTI-CLASS CLASSIFICATION $(K>2)$

$g_1(x)$
$g_2(x)$
$\vdots$
$g_K(x)$

$\underline{K=3}$

$y=1 \Leftarrow g_1(x) > g_2(x)$ & $g_1(x) > g_3(x)$

$y=2 \Leftarrow g_2(x) > g_1(x)$ & $g_2(x) > g_3(x)$

$y=3 \Leftarrow g_3(x) > g_1(x)$ & $g_3(x) > g_2(x)$

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

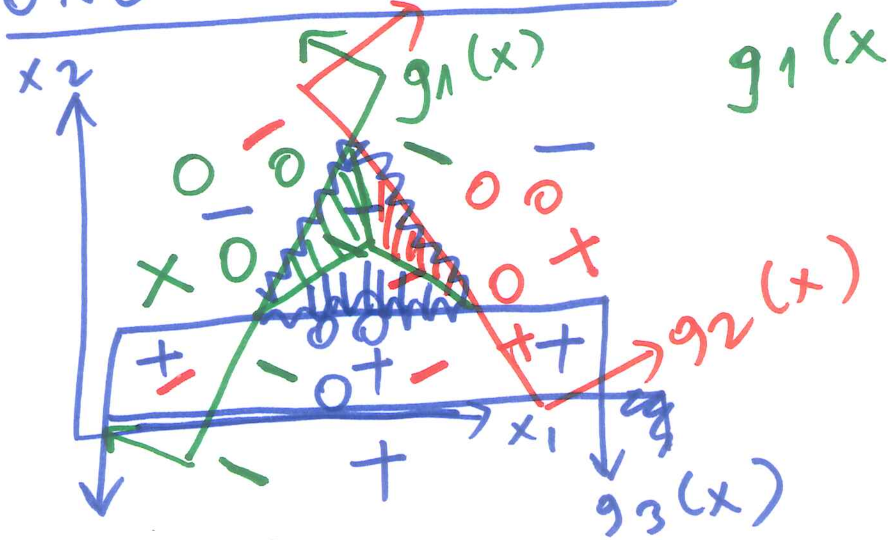
Error$(y, \hat{y})$

ONE-VERSUS-ALL (OVA)

−OTHERS

$g_1(x)$

o : 1st class
o : 2nd class
o : 3rd class



3-CLASS CLASSIFICATION PROBLEM

$\Downarrow$

3 2-CLASS CLASSIFICATION PROBLEMS

+	−	
ooo	oooooo	$\Rightarrow g_1(x)$
ooo	oooooo	$\Rightarrow g_2(x)$
ooo	oooooo	$\Rightarrow g_3(x)$

$g_1(x^*)$ $g_2(x^*)$ $g_3(x^*)$

K discriminants

ONE-VERSUS-OTHER (OVO)

+	−	
ooo	ooo	$g_{12}(x)$
ooo	ooo	$g_{13}(x)$
ooo	ooo	$g_{23}(x)$

5 4

$$10 = \frac{K(K-1)}{2} \text{ discriminants}$$

x^*

1	2	3	
✓			$g_{12}(x^*) > 0$
✓			$g_{13}(x^*) > 0$
		✓	$g_{23}(x^*) < 0$
$\boxed{2}$	0	1	# of wins

class #1

OVA versus OVO

$$K.(D+1) \qquad \underbrace{\frac{K.(K-1)}{2}.(D+1)}$$

$K > 3$ # of parameters in OVO is larger than that of OVA

$K = 2$

$p(y=1|x) = \delta$

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

$p(y=1|x) = \dfrac{p(x|y=1)\,p(y=1)}{p(x)}$

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

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

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

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

$$p(x|y=1) = \mathcal{N}(\mu_1, \Sigma) \qquad \Sigma_1 = \Sigma_2 = \Sigma \qquad \frac{\exp(a)}{\exp(b)} = \exp(a-b)$$

$$p(x|y=2) = \mathcal{N}(\mu_2, \Sigma)$$

$$\log\left[\frac{\cancel{(2\pi)^{-D/2}} \cdot \cancel{|\Sigma|^{-1/2}} \cdot \exp\left[\overbrace{-\tfrac{1}{2}(x-\mu_1)^T \cdot \Sigma^{-1} \cdot (x-\mu_1)}^{a}\right]}{\cancel{(2\pi)^{-D/2}} \cdot \cancel{|\Sigma|^{-1/2}} \cdot \exp\left[\underbrace{-\tfrac{1}{2}(x-\mu_2)^T \cdot \Sigma^{-1}(x-\mu_2)}_{b}\right]}\right] + \log\left[\frac{p(y=1)}{p(y=2)}\right]$$

Exercise #3

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

$$= \underline{\underline{w^T \cdot x + w_0}}$$

$$\Sigma \Rightarrow \hat{\Sigma} \qquad p(y=1) \Rightarrow p(\widehat{y=1})$$

$$\mu_1 \Rightarrow \hat{\mu}_1 \qquad p(y=2) \Rightarrow p(\widehat{y=2})$$

$$\mu_2 \Rightarrow \hat{\mu}_2$$