Linear Discrimination

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

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

$$c^* = \arg\max_{c=1}^{K} \underline{\underline{g_c(x)}}$$

$$g_c(x) = p(x \mid y=c)\,\Pr(y=c)$$

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

class conditional density

$\Downarrow$

$\mathcal{N}(\mu_c, \sigma_c^2)$

$\Downarrow$

$\hat{\mu}_c \hookrightarrow 1\times 1$, $\hat{\sigma}_c^2 \hookrightarrow 1\times 1$

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

class conditional density

$\Downarrow$

$\mathcal{N}(\mu_c, \Sigma_c)$

$\Downarrow$

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

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

$$\boxed{\begin{array}{l} \sum_{i=1}^{N} x_i \cdot 1 = \\ [x_1\ x_2 \dots x_N] \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \\ \texttt{np.sum( )} \end{array}}$$

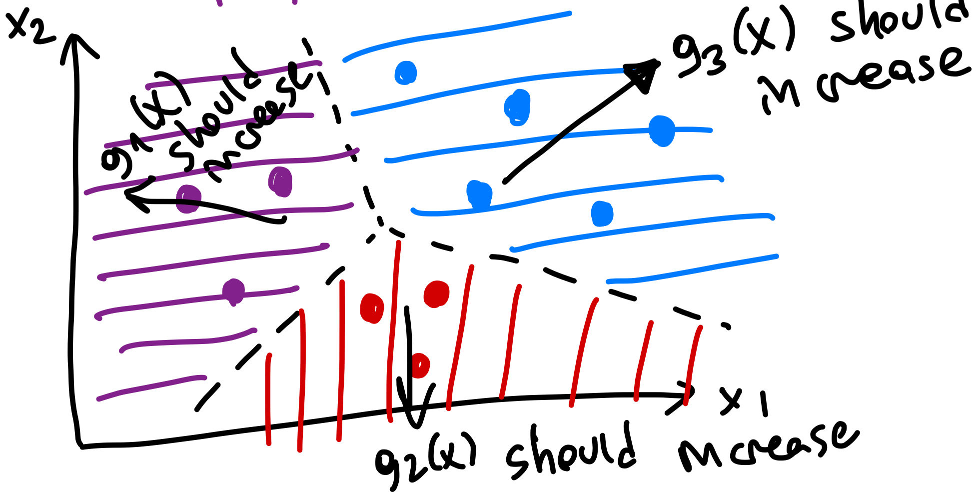
$$g_c(x \mid w_c, w_{c0}) = w_c^T . x + w_{c0}$$

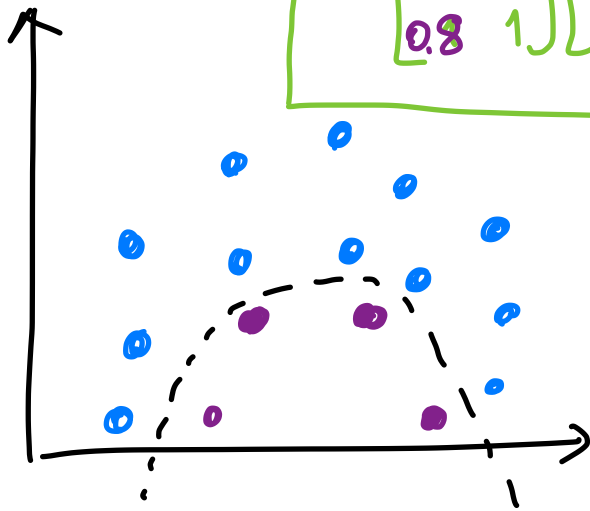
$$= [w_{c1}\ \ w_{c2}\ \ \dots\ \ w_{cD}]\begin{bmatrix}x_1\\x_2\\\vdots\\x_D\end{bmatrix} + w_{c0} = \sum_{d=1}^{D}(w_{cd}.x_d) + w_{c0}$$

$$a^2+2ab+b^2$$

$$[a,b]\begin{bmatrix}1 & 1.2\\0.8 & 1\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}$$

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





$$[a\ \ b]\begin{bmatrix}a+b\\a+b\end{bmatrix} =$$

$$a^2+ab+ba+b^2$$
$$= a^2+2ab+b^2$$

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

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

Binary Classification $(K=2)$

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

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

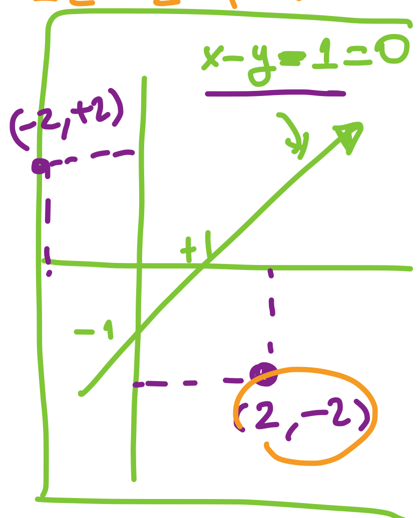
$$\text{if } g_1(x) - g_2(x) < 0 \quad \Rightarrow \hat{y} = 2$$

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

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

$$-x+y+1=0$$

$$-2-2-1<0$$



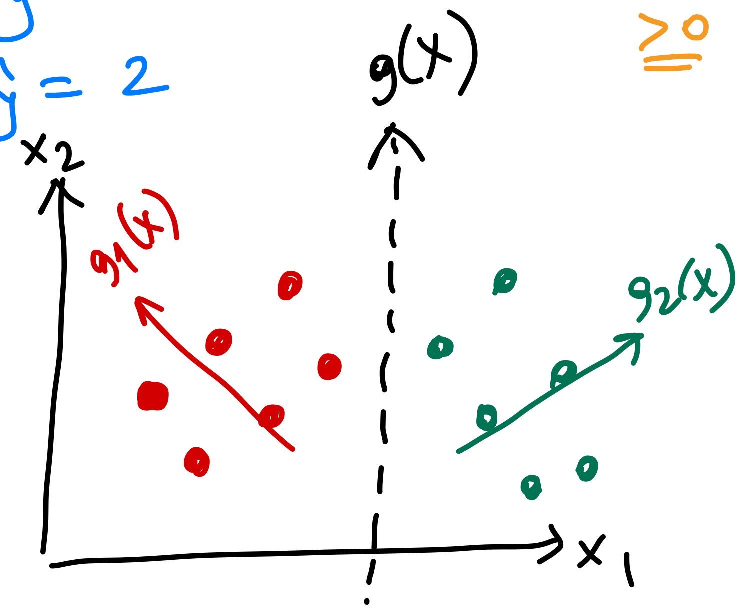
$$2-(-2)-1 \geq 0$$

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

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

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

$$= W^T . x + w_0$$



$x . a . x \rightarrow ax^2$ $\quad$ $x . b . x \rightarrow bx^2$

$$tr[x^T . W_1 x] = tr[W_1 . x x^T]$$

$$tr(ABC) = tr(BCA)$$

Multiclass Classification $(K>2)$

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

assume $\underline{K=3}$

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

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

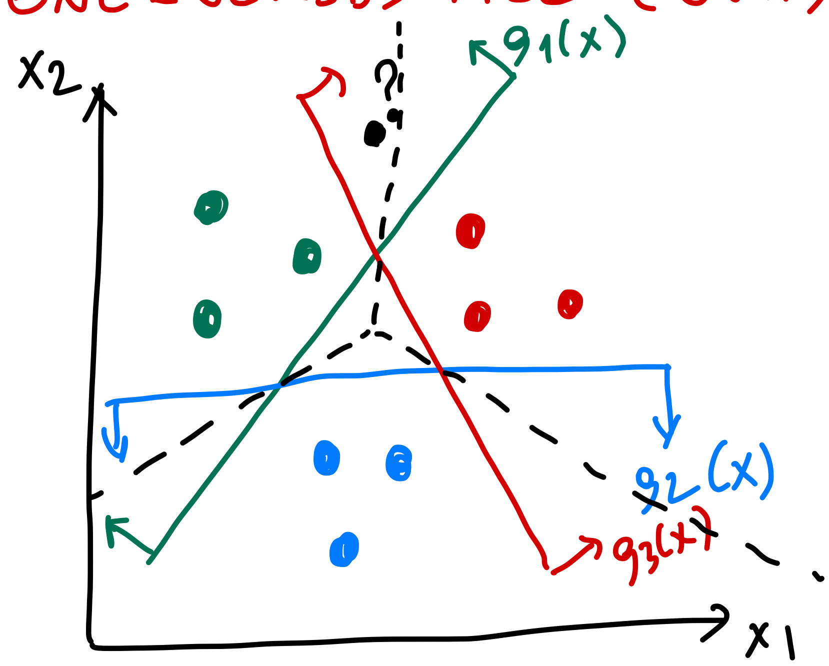
if $g_3(x) > g_1(x)$ & $g_3(x) > g_2(x) \Rightarrow \hat{y} = 3$

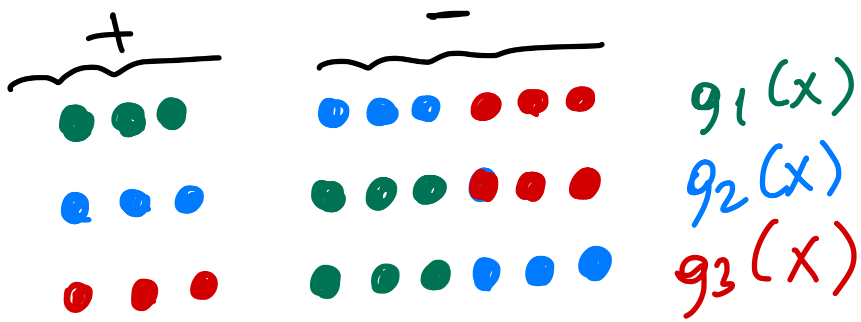
$$\boxed{\hat{y} = \arg\max_{c=1}^{K} g_c(x)}$$

$1x^2 + 0y^2 = 9$

$0x^2 + 1y^2 = 16$

ONE-VERSUS-ALL (OVA) APPROACH



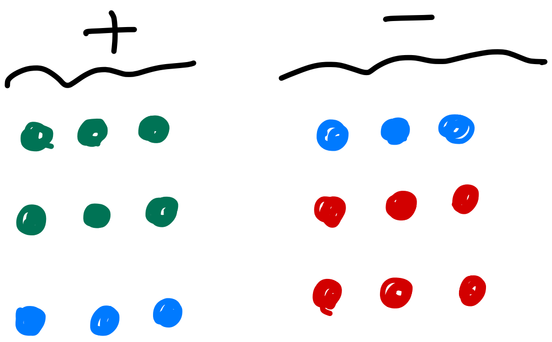


$\#$ of parameters $= K(D+1)$

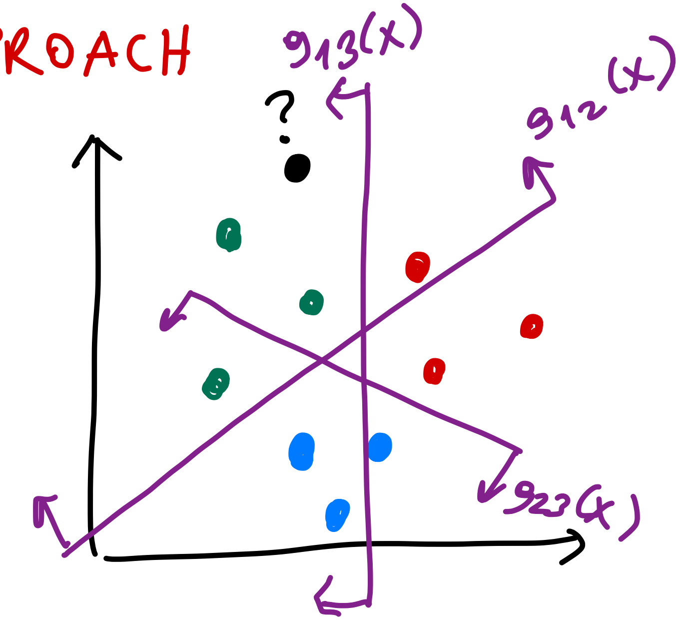
data set size for each problem $= N$

$\#$ of score functions $= K$

ONE-VERSUS-OTHER (OVO) APPROACH



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



$$\binom{K}{2} = \frac{K!}{(K-2)!\,2!} = \frac{K.(K-1)}{2}$$

?

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

overall winner

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

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

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

$K=2$

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

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

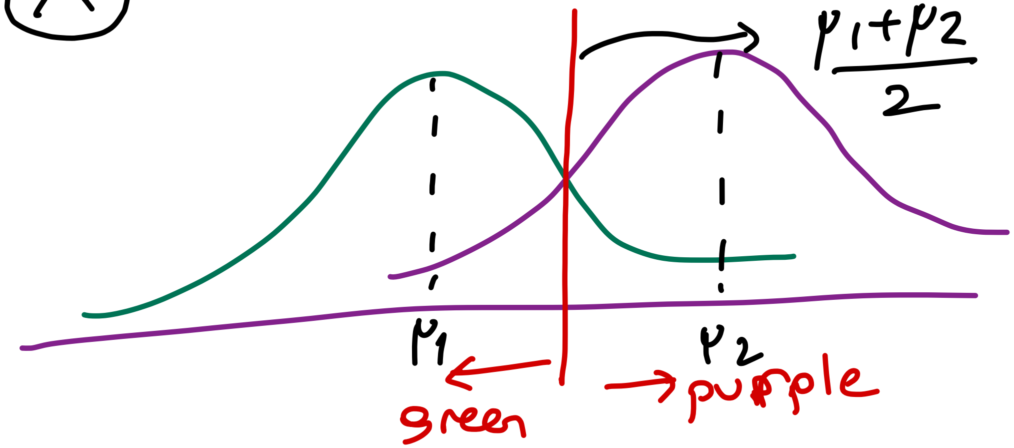
$$\log\left[\frac{\Pr(y=1 \mid x)}{\Pr(y=2 \mid x)}\right] = \log\left[\frac{\dfrac{p(x \mid y=1)\Pr(y=1)}{\cancel{p(x)}}}{\dfrac{p(x \mid y=2)\Pr(y=2)}{\cancel{p(x)}}}\right]$$

$$= \underbrace{\log\left[\frac{p(x \mid y=1)}{p(x \mid y=2)}\right]}_{A} + \underbrace{\log\left[\frac{\Pr(y=1)}{\Pr(y=2)}\right]}_{B}$$

$$p(x \mid y=1) \rightarrow N(x; \mu_1, \Sigma_1)$$
$$p(x \mid y=2) \rightarrow N(x; \mu_2, \Sigma_2)$$

assuming

$$\Sigma_1 = \Sigma_2 = \Sigma$$



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

$$= \underbrace{\log\left[\frac{\cancel{\frac{1}{\sqrt{(2\pi)^p|\Sigma|}}}\cdot \exp\left[-\frac{1}{2}(x-\mu_1)\Sigma^{-1}(x-\mu_1)\right]}{\cancel{\frac{1}{\sqrt{(2\pi)^p|\Sigma|}}}\cdot \exp\left[-\frac{1}{2}(x-\mu_2)\Sigma^{-1}(x-\mu_2)\right]}\right]}_{A} + \underbrace{\log\left[\frac{Pr(y=1)}{Pr(y=2)}\right]}_{B}$$

$$= -\frac{1}{2}(x-\mu_1)^T\cdot\Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)\Sigma^{-1}(x-\mu_2) + B$$

$$= \cancel{-\frac{1}{2}x^T\cdot\Sigma^{-1}\cdot x} + \mu_1^T\Sigma^{-1}\cdot x - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 \cancel{+\frac{1}{2}x^T\Sigma^{-1}\cdot x} - \mu_2^T\Sigma^{-1}x + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + B$$

$$= \underbrace{(\hat{\mu}_1-\hat{\mu}_2)^T\cdot\hat{\Sigma}^{-1}}_{w^T}\cdot x + \underbrace{\left[-\frac{1}{2}(\hat{\mu}_1+\hat{\mu}_2)^T\cdot\hat{\Sigma}^{-1}(\hat{\mu}_1-\hat{\mu}_2) + \log\left[\frac{\hat{Pr}(y=1)}{\hat{Pr}(y=2)}\right]\right]}_{w_0}$$