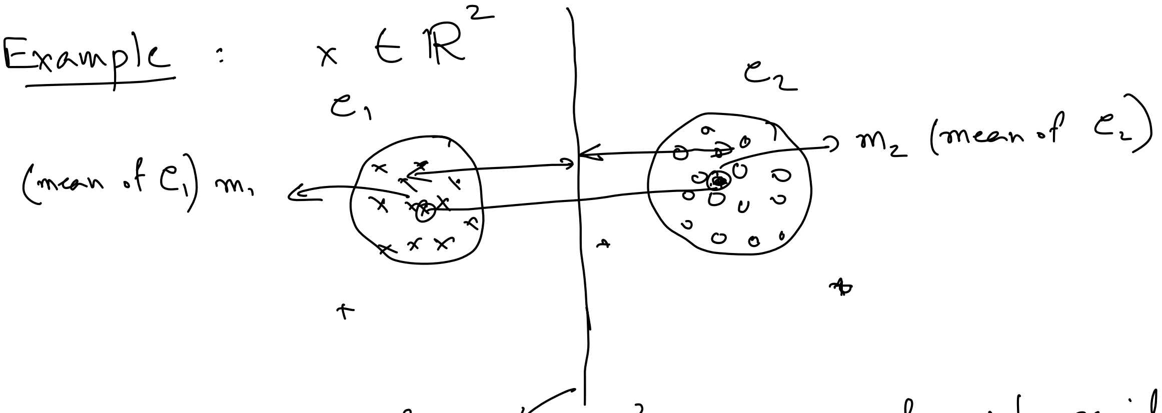
Classification

Training Set : $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$

Goal : Learn function f that predicts class label y of new (test) point x

Example : $x \in \mathbb{R}^2$



$$\|x - m_1\|_2^2 = \|x - m_2\|_2^2$$ – Locus of points equidistant to m_1 & m_2

$$(x-m_1)^T(x-m_1) = (x-m_2)^T(x-m_2)$$

$$\cancel{x^Tx} - 2x^Tm_1 + m_1^Tm_1 = \cancel{x^Tx} - 2x^Tm_2 + m_2^Tm_2$$

$$(m_2 - m_1)^Tx + \frac{1}{2}\left(\|m_1\|^2 - \|m_2\|^2\right) = 0$$

Decision surface

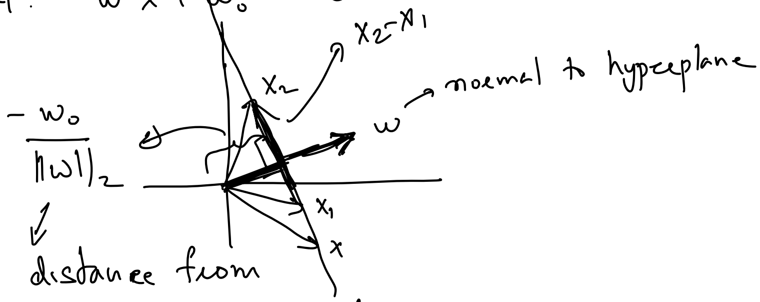
$$\boxed{y(x) = (m_2 - m_1)^Tx + \frac{1}{2}\left(\|m_1\|^2 - \|m_2\|^2\right)}$$ – has the form $\boxed{w^Tx + w_0}$ hyperplane

$w = m_2 - m_1$, $\quad \frac{1}{2}\left(\|m_1\|^2 - \|m_2\|^2\right)$

$y(x) > 0, \quad x \in C_2$

$y(x) < 0, \quad x \in C_1$

$H: \; w^Tx + w_0 = 0$



signed distance from origin to the hyperplane

If x lies on the hyperplane:

$$w^Tx + w_0 = 0$$

$$\boxed{\left(\frac{w}{\|w\|_2}\right)^T x = -\frac{w_0}{\|w\|_2}}$$

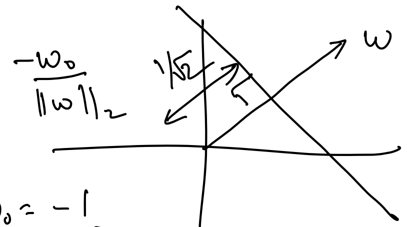
If x_1 & x_2 lie on the hyperplane :

$$w^Tx_1 + \cancel{w_0} = w^Tx_2 + \cancel{w_0}$$

$w^T(x_1 - x_2) = 0 \Rightarrow w$ is normal (perpendicular) to (the points on) hyperplane

Example : $x_1 + x_2 = 1$, $w = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. $w_0 = -1$

$$w^T x + w_0 = 0 \iff \begin{bmatrix} 1 & 1 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 1 = 0$$



$w_0 = -1$

$w = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \|w\|_2 = \sqrt{2}$

$-\frac{w_0}{\|w\|_2} = 1/\sqrt{2}$

C_1 , C_2

$p(C_1|x)$, $p(C_2|x)$

data likelihood

prior (class prior)

posterior

$$p(C_1|x) = \frac{p(x|C_1)\,p(C_1)}{p(x)}$$

$$p(C_2|x) = \frac{p(x|C_2)\,p(C_2)}{p(x)}$$

Bayes Rule

MAP rule

Maximum aposteriori probability : $\underset{i}{\operatorname{argmax}}\; p(C_i|x)$

$x \in \mathbb{R}^d$.

$$\boxed{\text{Gaussian Model} : p(x|C_i) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)}}$$

$|\Sigma| \to \det(\Sigma)$

MAP rule :

$$\underset{i}{\operatorname{argmax}}\; p(C_i|x)$$

$$= \underset{i}{\operatorname{argmax}}\; \log\left(p(C_i|x)\right)$$

Bayes rule

$$= \underset{i}{\operatorname{argmax}}\; \log\left(p(x|C_i)\,p(C_i)\right)$$

$$= \underset{i}{\operatorname{argmax}}\; \log p(x|C_i) + \log p(C_i)$$

Gaussian Model

$$\log p(x|C_i) = -\frac{d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)$$

Decision Surface : $-\log p(C_1|x) = -\log p(C_2|x)$

$$\cancel{\frac{d}{2}\log 2\pi} + \frac{1}{2}\log|\Sigma_1| + \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1) - \log p(C_1) =$$

$$\cancel{\frac{d}{2}\log 2\pi} + \frac{1}{2}\log|\Sigma_2| + \frac{1}{2}(x-\mu_2)^T \Sigma_2^{-1}(x-\mu_2) - \log p(C_2)$$

Case I : $\Sigma_1 = \Sigma_2 = I$

Decision Surface : $\frac{1}{2}\log|\Sigma_1| + \frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) - \log p(C_1) = \frac{1}{2}\log|\Sigma_2| + \frac{1}{2}(x-\mu)^T\Sigma_2^{-1}(x-\mu_2) - \log p(C_2)$

Plug in $\Sigma_1 = \Sigma_2 = I$

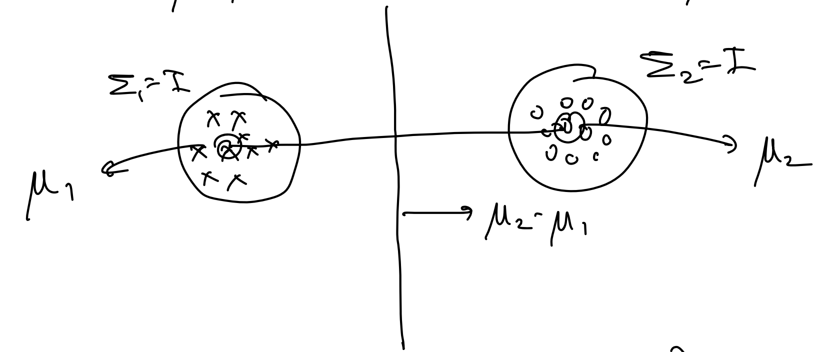
$$\frac{1}{2}(x-\mu_1)^T(x-\mu_1) - \log p(C_1) = \frac{1}{2}(x-\mu_2)^T(x-\mu_2) - \log p(C_2)$$

$$\frac{1}{2}(\cancel{x^Tx} - 2\mu_1^Tx + \mu_1^T\mu_1) - \log p(C_1) = \frac{1}{2}(\cancel{x^Tx} - 2\mu_2^Tx + \mu_2^T\mu_2) - \log p(C_2)$$

$$(\mu_2-\mu_1)^Tx + \frac{1}{2}(\|\mu_1\|^2 - \|\mu_2\|^2) - \log\frac{p(C_1)}{p(C_2)} = 0$$

Decision surface has the form : $w^Tx + w_0 = 0$

$$w = \mu_2 - \mu_1, \quad w_0 = \frac{1}{2}(\|\mu_1\|^2 - \|\mu_2\|^2) - \log\frac{p(C_1)}{p(C_2)}$$



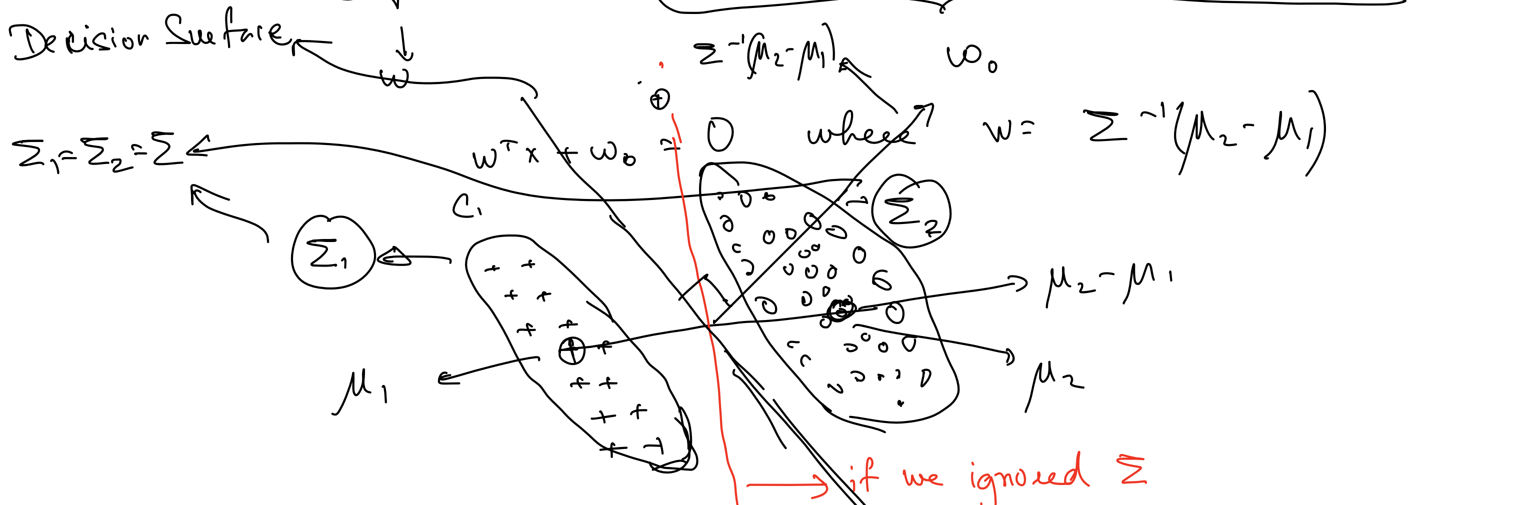
Case II : $\Sigma_1 = \Sigma_2 = \Sigma$ (but not necessarily I)

Decision Surface :

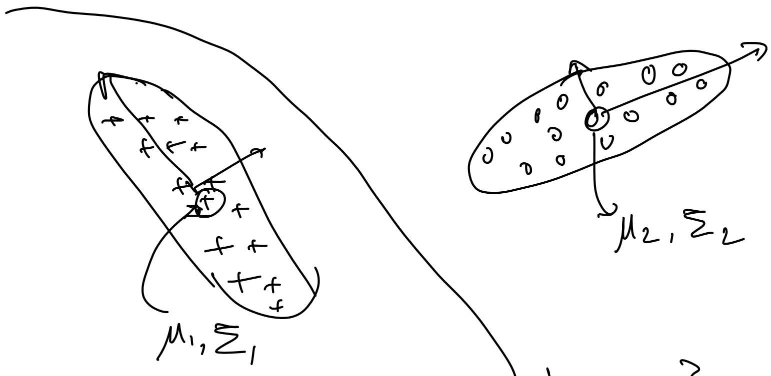
$$\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) - \log p(C_1) = \frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) - \log p(C_2)$$

$$\frac{1}{2}(x^T\Sigma^{-1}x - 2\mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1) - \log p(C_1) =$$
$$\frac{1}{2}(x^T\Sigma^{-1}x - 2\mu_2\Sigma^{-1}x + \mu_2^T\Sigma^{-1}\mu_2) - \log p(C_2)$$

$$\Rightarrow \underbrace{(\mu_2-\mu_1)^T\Sigma^{-1}}_{w}x + \underbrace{\frac{1}{2}(\mu_1+\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2) - \log\frac{p(C_1)}{p(C_2)}}_{w_0} = 0$$



Case III : Covariances Σ_1 and Σ_2 are different



Will the decision surface be linear?

Decision Surface :

$$\frac{1}{2}\log|\Sigma_1| + \frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) - \log p(C_1) = \frac{1}{2}\log|\Sigma_2| + \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2) - \log p(C_2)$$

Quadratic terms : $\frac{1}{2}x^T\Sigma_1^{-1}x$ & $\frac{1}{2}x^T\Sigma_2^{-1}x$

they will not cancel if $\Sigma_1 \neq \Sigma_2$

Decision Surface will be a general quadratic surface.