Extension to M-ary Bayes Classifier

$H_y : \underline{X} \sim p(\underline{x}|y)$, $y = 1, \ldots, M$

priors $\pi_y = P\{Y = y\}$, $y = 1, \ldots, M$

$\hat{y}_{Bayes} = f_{Bayes}(\underline{x}) =$

f_{Bayes} minimizes

Multivariate Gaussian Distribution

Recall : Gaussian random variable X with df

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Generalization to d-dimensional vector $\underline{X}$:

$$\phi(\underline{x}, \underline{\mu}, C) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|C|}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T C^{-1}(\underline{x}-\underline{\mu})}$$

$\underline{\mu} =$, $C =$

$\mu_l =$, $C_{l,m} =$

Compact Notation :

M-ary Classification with Multivariate Gaussians

$$H_y : p(\underline{x} \mid y) = \phi(\underline{x}, \underline{\mu}_y, C_y)$$

$$\hat{y}_{Bayes} = f_{Bayes}(\underline{x}) = \arg\max_y \left(\ln \pi_y + \ln p(\underline{x} \mid y) \right)$$

$$p(\underline{x} \mid y) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|C_y|}} e^{-\frac{1}{2}(\underline{x} - \underline{\mu}_y)^T C_y^{-1} (\underline{x} - \underline{\mu}_y)}$$

$$\ln p(\underline{x} \mid y) =$$

$$\hat{y}_{Bayes} = \arg\max_y$$

If C_y's are different,

If $C_y = C$, for all y, then

$$\frac{1}{2}(\underline{x} - \underline{\mu}_y)^T C_y^{-1} (\underline{x} - \underline{\mu}_y) = \frac{1}{2}\left(\underline{x}^T C^{-1} \underline{x} - \underline{\mu}_y^T C^{-1} \underline{x} - \underline{x}^T C^{-1} \underline{\mu}_y + \underline{\mu}_y^T C^{-1} \underline{\mu}_y \right)$$

$$\Rightarrow \hat{y}_{Bayes} = \arg\max_y$$

Linear Discriminant Analysis (LDA) Classifier

- Assuming $C_y = C$, for all y,

$$f_{Bayes}(x) = \arg\max_{y} \left[\ln \pi_y + \underline{\mu}_y^T C^{-1} \underline{x} - \frac{1}{2} \underline{\mu}_y^T C^{-1} \underline{\mu}_y \right]$$

- LDA approach replaces $\{\pi_\ell, \underline{\mu}_\ell\}_{\ell=1}^{M}$ and C by estimates and that are obtained from

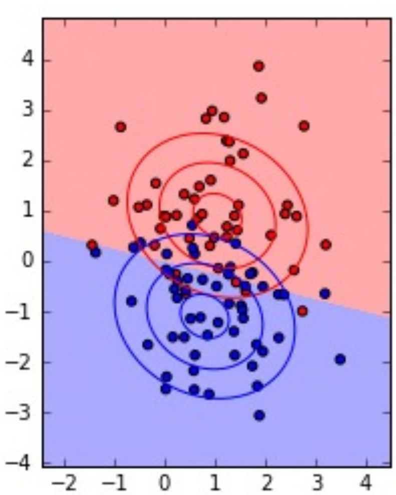
<u>Estimation</u> Let N_ℓ = # training data with class label ℓ

$\hat{\pi}_\ell =$, $\hat{\mu}_\ell =$

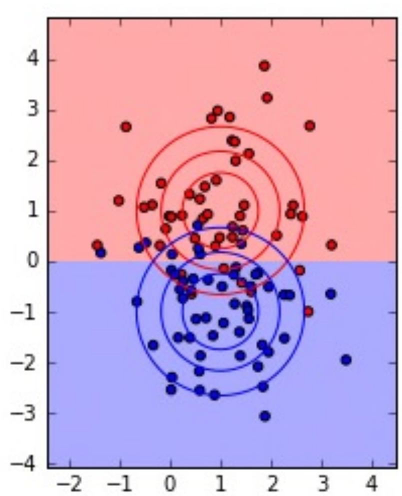
$\hat{C} =$

Let $\delta_\ell(\underline{x}) = \ln \hat{\pi}_\ell + \hat{\underline{\mu}}_\ell^T \hat{C}^{-1} \underline{x} - \frac{1}{2} \hat{\underline{\mu}}_\ell \hat{C}^{-1} \hat{\underline{\mu}}_\ell$

Then $\hat{y}_{LDA} = f_{LDA}(\underline{x}) =$



LDA



Bayes

Linear Classifiers

- Linear classifiers (e.g. LDA) are classifiers for which the objective $\delta_\ell(\underline{x})$ is linear

$$\delta_\ell(\underline{x}) =$$

For LDA, $\underline{a}_\ell^T =$, $b_\ell =$

$$\hat{y}_{lin} = f_{lin}(\underline{x}) =$$

- Decision between classes ℓ and m:

$$\underline{a}_\ell^T \underline{x} + b_\ell \underset{m}{\overset{\ell}{\gtrless}} \underline{a}_m^T \underline{x} + b_m$$

$$\equiv$$

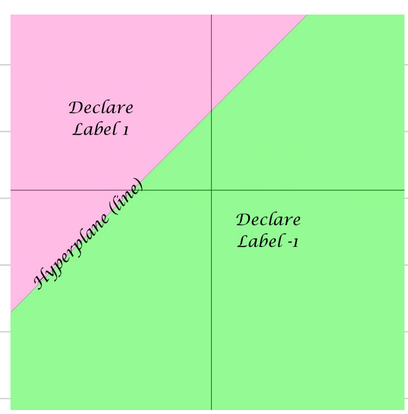
- Define $g_{\ell m}(\underline{x}) =$

Then

- $g_{\ell m}(\underline{x}) = 0$ defines a

Binary Linear Classifier

$$\hat{y}_{lin} = f_{lin}(x) = \begin{cases} 1 & \text{if } \underline{w}^T \underline{x} \geq \beta \\ -1 & \text{if } \underline{w}^T \underline{x} < \beta \end{cases}$$



Example $\quad M = 3, \; d = 2, \quad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\delta_1(\underline{x}) = x_1 + 2x_2 + 1$$
$$\delta_2(\underline{x}) = 2x_1 - x_2 + 2$$
$$\delta_3(\underline{x}) = -x_1 + 2x_2 + 1$$

Linear discriminant functions :

Suppose we want to classify $\underline{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

1. Using $\delta_e(\underline{x})$:

2. Using $g_{em}(\underline{x})$: