# Classification

**Goal:** Predict $Y \in G$ from $\vec{x}$

$$\underbrace{G}_{\text{discrete set}}$$

## Linear Discriminant Analysis (LDA)

We will assume

$$X \mid Y_i = 1 \sim f_1(x)$$

$$X \mid Y = 2 \sim f_2(x)$$

$$\vdots$$

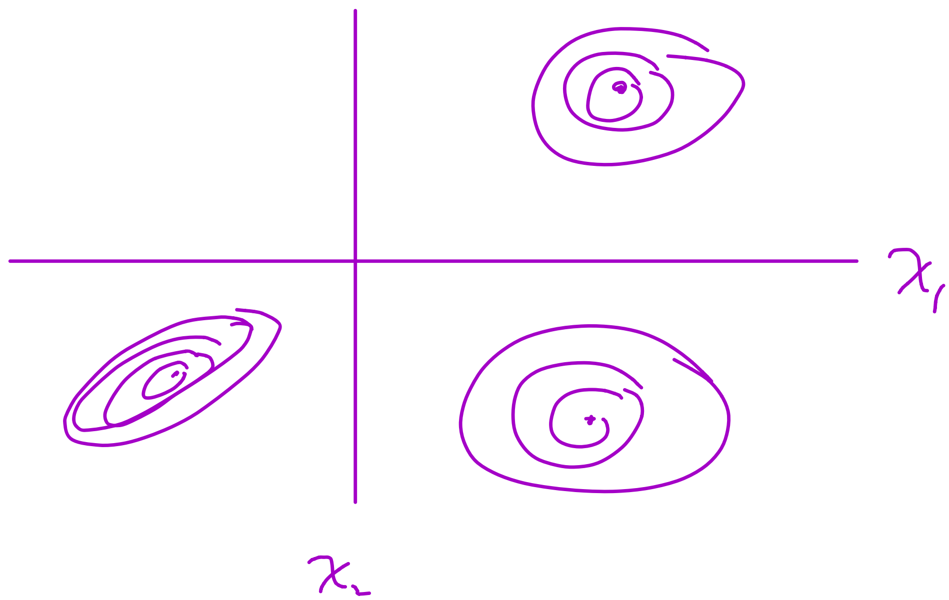
$$X \mid Y = K \sim f_K(x)$$

$$\mathbb{P}(Y = k \mid X = x) = \frac{\mathbb{P}(X = x \mid Y = k)\,\mathbb{P}(Y = k)}{\mathbb{P}(X = x)}$$

$$= \frac{\mathbb{P}(X = x \mid Y = k)\,\mathbb{P}(Y = k)}{\sum^{K} \mathbb{P}(X = x \mid Y = \ell)\,\mathbb{P}(Y = \ell)}$$

$l=1$

$$= \frac{f_k(x)\,\pi_k}{\sum_{l=1}^{K} f_l(x)\,\pi_l}$$




Classification rule: $\hat{y} = \underset{k}{\operatorname{argmax}}\ f_k(x)\,\pi_k$

Consider the likelihood ratio

$$\frac{P(y=k \mid X=x)}{P(y=l \mid X=x)} = \frac{f_k(x)\,\pi_k}{\quad}$$

$\mathbb{P}(Y=l \mid X=x) \qquad f_l(x)\,\pi_l$

and the log-odds follows as

$$\log \frac{\mathbb{P}(Y=k \mid X=x)}{\mathbb{P}(Y=l \mid X=x)} = \log \frac{f_k(x)\,\pi_k}{f_l(x)\,\pi_l}$$

$$= \log \frac{f_k(x)}{f_l(x)} + \log \frac{\pi_k}{\pi_l}$$

$=$ Assuming each density has a common variance in the MVN

$$= \log \frac{\pi_k}{\pi_l} - \frac{1}{2}(\mu_k+\mu_l)^T \Sigma^{-1} (\mu_k-\mu_l)$$

$$+ x^T \Sigma^{-1} (\mu_k-\mu_l)$$

* linear in $X$

* Descion boundary

$$\{x : \mathbb{P}(G=k \mid X=x) = \mathbb{P}(G=\ell \mid X=x)\}$$

is linear in $X$ and in

$p$ dimensional space.

Define the <u>linear discriminant functions</u>

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k$$

is equivalent to the decision

rule

$$\hat{Y} = \underset{k}{\operatorname{argmax}}\ \delta_k(x)$$

## Connection to linear models

$$\log \frac{\mathbb{P}(Y=k \mid X=x)}{\mathbb{P}(Y=\ell \mid X=x)} = \underbrace{\log \frac{\pi_k}{\pi_\ell} - \frac{1}{2}(\mu_k+\mu_\ell)^T \Sigma^{-1}(\mu_k-\mu_\ell)}_{\beta_0}$$

$$+ x^T \underbrace{\Sigma^{-1}(\mu_k-\mu_\ell)}_{\beta_1}$$

So the decision rule is based on this linear function