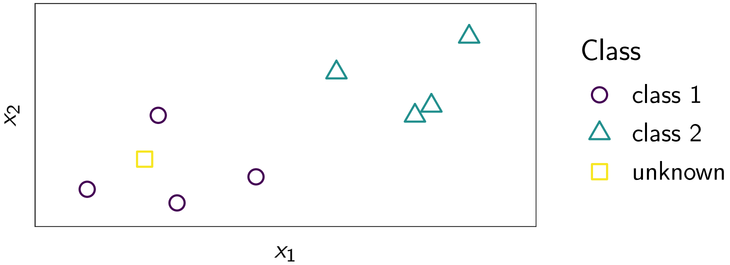
Discriminant Analysis as an Example of Classification Techniques for Compositional Data

Manuel Pfeuffer

Seminar: Compositional Data Analysis
Humboldt–Universität zu Berlin

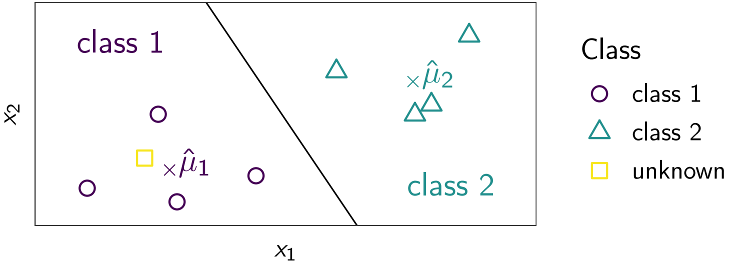
When do we need classification techniques?



Classification Problem

Given n observations $x_i = (x_{i1}, \ldots, x_{ik})$ with known class labels $y_i \in \{1, \ldots, g\}$, predict the class of an unlabelled observation.

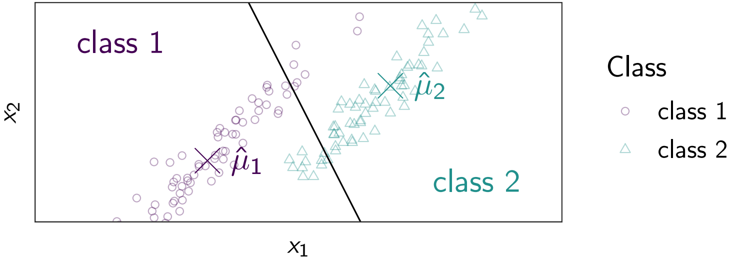
A simple approach to classification



Discriminant Function and Classifier

$$\delta_j(x) = \|x - \hat{\mu}_j\|, \qquad D(x) = \begin{cases} \text{"class 1"}, & \text{if } \delta_1 < \delta_2 \\ \text{"class 2"}, & \text{otherwise} \end{cases}$$

A simple approach to classification



Problem:

- $\delta(x)$ only takes into account the μ_j' s
- This is problematic for correlated data

$\Rightarrow$ We should also take into account the Σ_j's!

Bayes Discriminant Rule

Assumption

Class populations characterized by a density function f_j.

$\rightarrow$ Usually f_j assumed mutltivariate normal with (μ_j, Σ_j).

Conditional probability that observation x comes from class k:

$$P(G = k|x) = \frac{f_k(x)p_k}{\sum_{j=1}^{g} f_j(z)p_j} \quad \text{(Bayes Theorem)}$$

To compare conditional probabilities:

$$\delta_k(x) = \ln f_k(x) + \ln p_k \quad (\propto P(G = k|x))$$

$\Rightarrow$ Classify as class with the highest $\delta_k(x)$ (probability)!

Bayes Discriminant Rule: QDA and LDA

Plugging in the multivariate normal density for f_k:

$$\ln f_k(x) \propto -\frac{1}{2} ln|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)$$

Bayes Quadratic Discriminant Function

$$\delta_k^{QDA}(x) = -\frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) + \ln p_k$$

- "Quadratic" because $\delta_k^{QDA}(x)$ is quadratic in x.
- Linear discriminant analysis assumes $\Sigma_1 = \cdots = \Sigma_g$.
  $\rightarrow$ The resulting $\delta_k^{LDA}(x)$ is linear in x.
- Bayes Discriminant Analysis gives probabilistic output.

Bayes Discriminant Rule: Visualization

Note: Estimate μ_j and Σ_j by group mean and empirical covariance.

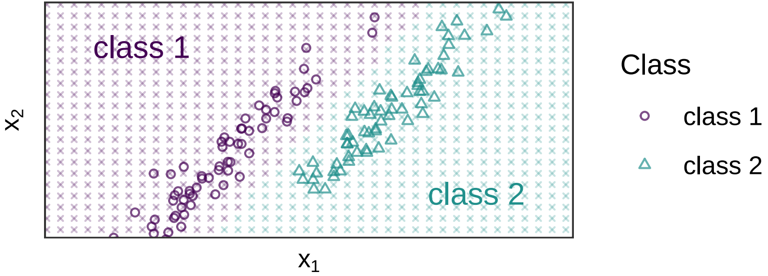


Figure: LDA on simulated data using `lda()` from the `MASS` package.
[To Check: Does that function really implement Bayes LDA?]

Fisher Discriminant Rule

Goal: Find a projection direction $a \in \mathbb{R}^k$ that:

1. Maximizes spread between the group means.

$$\mathbf{S_B} = \frac{1}{n} \sum_{1=j}^{g} n_j (\mu_j - \mu)(\mu_j - \mu)^T$$

2. Minimizes the variance within the groups.

$$\mathbf{S_W} = \frac{1}{n} \sum_{1=j}^{g} n_j \Sigma_j$$

Maximization Problem

$$\max \frac{a^T \mathbf{S_B} a}{a^T \mathbf{S_W} a} \quad \text{for } a \in \mathbb{R}^k, a \neq 0$$

Fisher Discriminant Rule: Visualization

Summary