Section 4I. Univariate LDA

Statistics for Data Science

Victor M. Preciado, PhD MIT EECS
Dept of Electrical & Systems Engineering
University of Pennsylvania
preciado@seas.upenn.edu

Univariate LDA

- **Univariate case** $(p = 1)$:
  - For $p = 1$, the Gaussian density has the form

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2}$$

  - Assuming identical variances $\sigma_k = \sigma$, we have that

$$\begin{aligned} \arg\max_k f_k(x)\,\pi_k &= \arg\max_k \log\left(f_k(x)\,\pi_k\right) = \cdots \\ &= \arg\max_k \left(\frac{\mu_k x}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log \pi_k\right) = \arg\max_k \delta_k(x) \end{aligned}$$

    where $\delta_k(x)$ is a linear function of x
  - In practice, we are not explicitly given μ_k, σ, and π_k, but a dataset...

Univariate LDA (cont.)

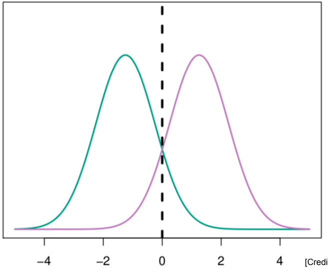
- We can estimate the parameters μ_k, σ, and π_k from a dataset, using the following formulas:

$$
\begin{aligned}
\widehat{\pi}_k &= \frac{N_k}{N} \\
\widehat{\mu}_k &= \frac{1}{N_k} \sum_{i:\ y_i=k} x_i \\
\widehat{\sigma}^2 &= \frac{1}{N-K} \sum_{k=1}^{K} \sum_{i:\ y_i=k} (x_i - \widehat{\mu}_k)^2 \\
&= \sum_{k=1}^{K} \frac{N_k - 1}{N-K} \widehat{\sigma}_k^2
\end{aligned}
$$

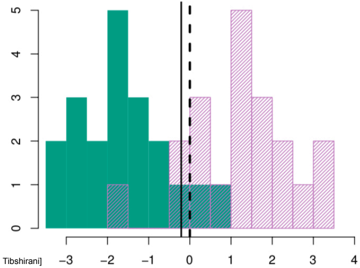
where N_k is the number of samples in the class k and $\widehat{\sigma}_k^2 = \frac{1}{N_k-1} \sum_{i:\ y_i=k} (x_i - \widehat{\mu}_k)^2$

Univariate LDA Example

Consider two Gaussians with $\mu_1 = -1.5$, $\mu_2 = 1.5$, $\pi_1 = \pi_2 = 0.5$ and $\sigma^2 = 1$



[Credit: Hastie and Tibshirani]

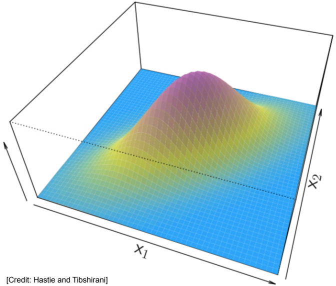


Multivariate LDA

- Multivariate LDA ($p > 1$): The Gaussian density takes the form

$$f_k(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_k)^T\Sigma^{-1}(\mathbf{x}-\mu_k)}$$

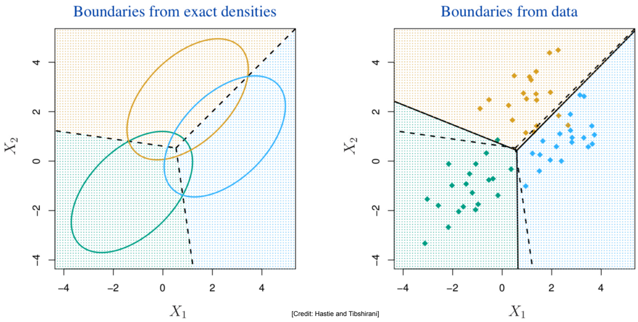
  where $\mathbf{x}$ and μ_k are p-dimensional vectors and Σ is a $p \times p$ matrix (and $|\Sigma|$ is its determinant)



[Credit: Hastie and Tibshirani]

Multivariate LDA: Numerical example

- Consider an example with $p = 2$ and $K = 3$ with $\pi_k = 1/3$:



[Credit: Hastie and Tibshirani]