M339G: March 10th, 2025.

Linear Discriminant Analysis w/ $p=1$.

Fisher 1936.

Goal: Classify observations into one of K classes ($K \geq 2$), i.e., figure out

$$\boxed{p_k(x) := \mathbb{P}[Y=k \mid X=x]}$$

posterior probability

Environment:

- π_k ... prior probability that a randomly chosen observation falls into category $k=1,\dots,K$
- $f_k(x)$ ... density function of X for observations from class $k=1..K$

choice, model

$f_k(x)dx$ ... the probability that X falls in $(x, x+dx)$ for points from class $k=1..K$

Then,

$$\mathbb{P}[Y=k \mid X=x] = \frac{\mathbb{P}[Y=k \text{ and } X=x]}{\mathbb{P}[X=x]} = \frac{\overbrace{\mathbb{P}[X=x \mid Y=k]}^{f_k(x)} \cdot \overbrace{\mathbb{P}[Y=k]}^{\text{"}\pi_k\text{"}}}{\mathbb{P}[X=x]}$$

$X \in (x, x+dx)$

Bayes Theorem

The Law of Total Probability

And,

$$\boxed{p_k(x) = \frac{f_k(x)\pi_k}{\sum_{\ell=1}^{K} f_\ell(x)\cdot \pi_\ell}}$$

$\Rightarrow$ classify into

$$\boxed{k = \operatorname{argmax}_{\ell=1..K} (p_\ell(x))}$$

The Specifics of LDA ($p=1$).

The choice: f_k are normal densities for each $k=1..K$, i.e.,

$$\boxed{f_k(x) = \frac{1}{\sigma_k\sqrt{2\pi}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}} \quad \text{for } k=1..K$$

w/ μ_k the mean and σ_k the std deviation for class k

Additional Assumption: Homogeneity: $\boxed{\sigma_1 = \dots = \sigma_K = \sigma}$

We now return to the posterior probability, i.e.,

$$p_k(x) = \frac{\pi_k f_k(x)}{\boxed{\sum_{\ell=1}^{K} \pi_\ell f_\ell(x)}}$$

Remember: We're looking for the k for which the above is MAXIMUM

Since all $p_k(x)$ have the same denominator, it's sufficient to find the k such that

$$\pi_k f_k(x) \longrightarrow \max$$

Because $\ln(\cdot)$ is increasing, the above is equivalent to

$$\ln(\pi_k) + \ln(f_k(x)) \longrightarrow \max$$

$$\Longleftrightarrow$$

$$\ln(\pi_k) + \ln\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_k)^2}{2\sigma^2}}\right) \longrightarrow \max$$

$$\Longleftrightarrow$$

$$\ln(\pi_k) - \underbrace{\boxed{\ln(\sigma\sqrt{2\pi})}}_{\text{constant in terms of } k} - \frac{(x-\mu_k)^2}{2\sigma^2} \longrightarrow \max$$

$$\Longleftrightarrow$$

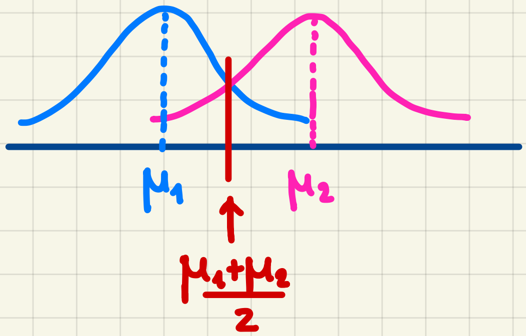
$$\ln(\pi_k) \boxed{- \frac{x^2}{2\sigma^2}} + \frac{\cancel{2}x\mu_k}{\cancel{2}\sigma^2} - \frac{\mu_k^2}{2\sigma^2} \longrightarrow \max$$

$$\delta_k(x) := \ln(\pi_k) + \frac{\mu_k}{\sigma^2}x - \frac{\mu_k^2}{2\sigma^2} \longrightarrow \max$$

These are called DISCRIMINANT (SCORES) and they are LINEAR in x.

Special Case:

$$\boxed{\underline{K}=2,\ \pi_1=\pi_2=\frac{1}{2}}$$



$$\downarrow$$

$$\frac{\mu_k}{\cancel{\sigma^2}}x-\frac{\mu_k^2}{2\cancel{\sigma^2}} \longrightarrow \max$$

$$\mu_k x-\frac{\mu_k^2}{2} \longrightarrow \max$$

IF

$\mu_1 x-\frac{\mu_1^2}{2} > \mu_2 x-\frac{\mu_2^2}{2}$, then classify as "1"

$$\Longleftrightarrow$$

$$\boxed{(\mu_1-\mu_2)}x > \frac{1}{2}(\mu_1^2-\mu_2^2)=\frac{1}{2}\boxed{(\mu_1-\mu_2)}(\mu_1+\mu_2)$$

Boundary always $\frac{\mu_1+\mu_2}{2}$