M339G: March 18th, 2024.

Linear Discrimant Analysis w/ $p=1$. (Fisher in 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 $k=1..K$
- $f_k(x)$ .... density function of X for observations from class $k=1..K$

choice

model

$\boxed{f_k(x)dx}$ ... the probability that X falls is $(x, x+dx)$ for points in class k

Then,

$$\mathbb{P}[Y=k \mid X=x] = \frac{\mathbb{P}[Y=k \text{ and } \overset{\in(x,x+dx)}{X=x}]}{\mathbb{P}[X=x]}$$

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

conditional density from model

Bayes Thm

$$= \frac{\boxed{\mathbb{P}[X=x \mid Y=k]} \cdot \boxed{\mathbb{P}[Y=k]}}{\boxed{\mathbb{P}[X=x]}}$$

π_k

The Law of Total Probability

$$\boxed{p_k(x) = \frac{\pi_k \cdot f_k(x)}{\sum_{j=1}^{K} \pi_j f_j(x)}}$$

$\Rightarrow$ classify into $k = \underset{j=1..K}{\operatorname{argmax}}\,(p_j(x))$

Linear Discriminant Analysis (LDA)

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

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

for $k=1..K$ ✓

w/ μ_k and σ_k being the mean and the sd deviation for the k^{th} class

Additional Assumption: Homogeneity $\sigma_1 = \cdots = \sigma_K = \sigma$

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

$$p_k(x) = \frac{\pi_k \cdot f_k(x)}{\sum_{j=1}^{K} \pi_j f_j(x)}$$

Remember: we're looking for k for which the above is maximal.

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

$$\pi_k f_k(x) \longrightarrow \text{maximal}$$

Because ln is increasing, this is equivalent to:

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

$$\Longleftrightarrow$$

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

$$\Longleftrightarrow$$

$$\ln(\pi_k) - \underbrace{\ln(\sigma\sqrt{2\pi})}_{\text{const.}} - \frac{(x-\mu_k)^2}{2\sigma^2} \longrightarrow \max$$

$$\Longleftrightarrow$$

$$\ln(\pi_k) \underbrace{- \frac{x^2}{2\sigma^2}}_{\text{doesn't depend on } k} + \frac{\cancel{2} x \mu_k}{\cancel{2}\sigma^2} - \frac{\mu_k^2}{2\sigma^2} \longrightarrow \max$$

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

These are called DISCRIMINANTS and they're $\boxed{\text{LINEAR}}$ in x.

Special Case: $K=2$, $\pi_1 = \pi_2 = \frac{1}{2}$

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

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

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

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

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

Bivariate Normal Variables. (Based on Pitman's "Probability")

Recall: In 1-D, the standard normal density is

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{for } z \in \mathbb{R}$$

In 2-D, we start w/ X and Y that are independent and standard normal ($N(0,1)$), then, their joint density, i.e., the density of the pair (X, Y) is

$$f_{XY}(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} \quad \text{for all } (x,y) \in \mathbb{R}^2$$