Aprendizaje de máquina

Análisis discriminante lineal

Ingeniería

Análisis discriminante

- Aquí el enfoque es modelar la distribución de X en cada una de las clases por separado y luego usar el teorema de Bayes para invertir las cosas y obtener $\mathbb{P}(Y|X)$.
- Cuando usamos distribuciones normales (gaussianas) para cada clase, esto conduce a un análisis discriminante lineal o cuadrático.
- Sin embargo, este enfoque es bastante general y también se pueden utilizar otras distribuciones. Nos centraremos en las distribuciones normales.

Teorema de Bayes

Probabilidad a-posteriori

Verosimilitud

Probabilidad a-priori

$$\mathbb{P}(Y = k|X = x) = \frac{\mathbb{P}(X = x|Y = k) \cdot \mathbb{P}(Y = k)}{\mathbb{P}(X = x)}$$

Factor de normalización

$$\mathbb{P}(X = x) = \sum_{l=1}^{K} \mathbb{P}(X = x|Y = l) \cdot \mathbb{P}(Y = l)$$

Teorema de Bayes

Probabilidad a-posteriori

Verosimilitud

Probabilidad a-priori

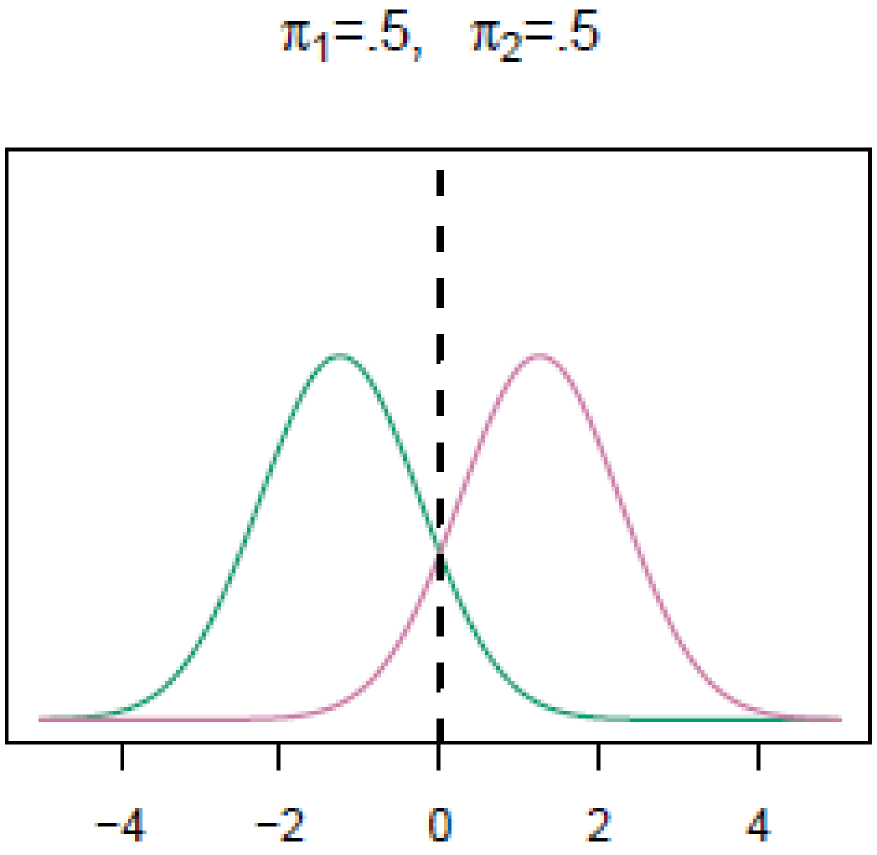
$$\mathbb{P}(Y = k | X = x) = \frac{f_k(x) \cdot \pi_k}{\sum_{l=1}^{K} f_l(x) \cdot \pi_l}$$

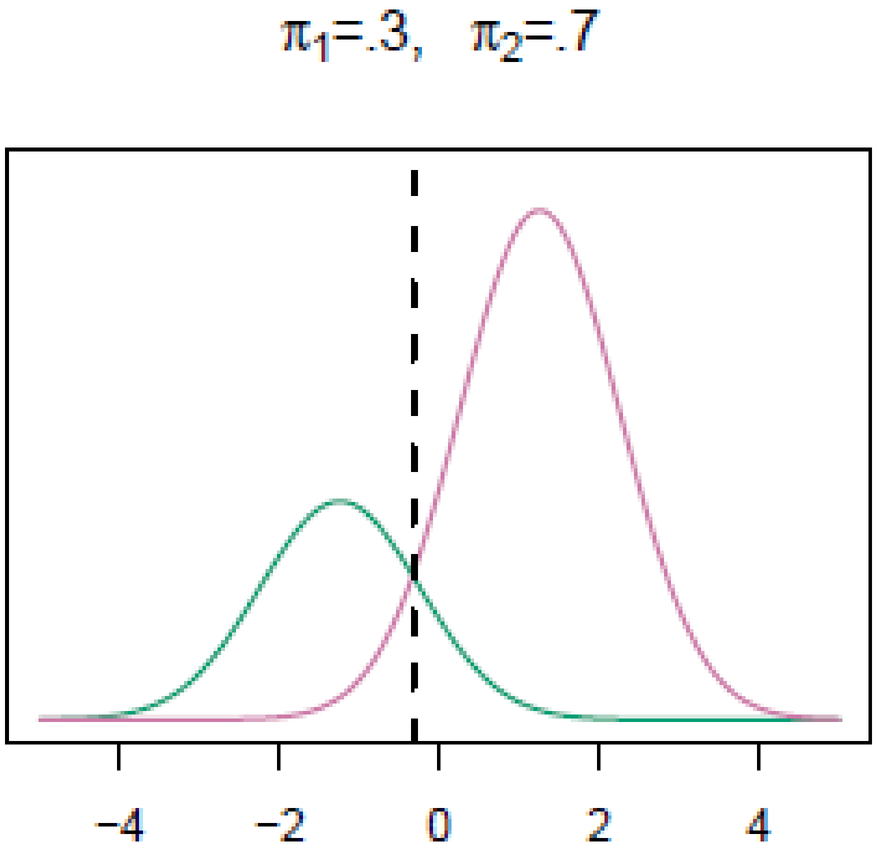
Factor de normalización

$f_k(x) = \mathbb{P}(X = x | Y = k)$ es la densidad de X en la clase k.

π_k es la probabilidad marginal o a-priori de la clase k.

Clasificamos utilizando la mayor densidad





Análisis discriminante lineal para $\mathrm{X} \in \mathbb{R}$

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

- Donde μ_k es la media y σ_k^2 la varianza en la clase k. Vamos a asumir que todas las $\sigma_k = \sigma$ son las mismas.

Análisis discriminante lineal para $\mathrm{X} \in \mathbb{R}$

- Como debemos seleccionar el máximo, podemos ignorar el factor de normalización, ya que este es el mismo para todas las clases.

$$\mathbb{P}(Y = k|X = x) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2} \cdot \pi_k$$

$$\ln \mathbb{P}(Y = k|X = x) = \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2}\left(\frac{x - \mu_k}{\sigma}\right)^2 + \ln \pi_k$$

Análisis discriminante lineal para $\mathrm{X} \in \mathbb{R}$

$$\ln \mathbb{P}(Y = k|X = x) = \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2}\left[x^2 - 2x\mu_k + \mu_k^2\right] + \ln \pi_k$$

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

Estimamos los parámetros

$$\hat{\pi}_k = \frac{n_k}{n}$$

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i:y_i=k} x_i$$

$$\hat{\sigma}^2 = \sum_{k=1}^{K} \frac{n_k - 1}{n - K} \frac{1}{n_k - 1} \sum_{i:y_i=k} (x_i - \hat{\mu}_k)^2$$

Análisis discriminante lineal para $\mathrm{X} \in \mathbb{R}^p$ donde p>1

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

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

Bayes básico

- La matriz de varianzas y covarianzas se asume diagonal debido a supuesto de independencia entre las características

$$\delta_k(x) \propto \ln\left[\pi_k \prod_{j=1}^{p} f_{kj}(x_j)\right]$$

$$= -\frac{1}{2}\sum_{j=i}^{p}\left[\frac{\left(x_j - \mu_{kj}\right)^2}{\sigma_{kj}^2} + \ln \sigma_{kj}^2\right] + \ln \pi_k$$

¿Cuándo usarlo?

- LDA es útil cuando n es pequeño o las clases están bien separadas y las suposiciones de normalidad son razonables. También cuando $K > 2$.
- Bayes ingenuo es útil cuando p es muy grande.