Estadística Multivariada

– Ejercicios Jueves 14 de Marzo del 2024 –

Israel García Ramírez – 211179

Ejercicio 1. Sea $p_k(x) = \mathbb{P}(y = k \mid x)$ con

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

(1) Calcula $p_k(x)$ suponiendo homogeneidad.

(2) Sea $K = 2$ y $\pi_1 = \pi_2 = 1/2$. Verifica que la frontera de decisión se encuentra en

$$\frac{\mu_1 + \mu_2}{2}.$$

(1)

$$p_K(x) = \mathbb{P}(y=K \mid x) = \frac{f_i(x)\pi_i}{\sum_{K=1}^{n} f_K(x)\pi_K}$$

$$= \frac{\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma}\right)^2\right\}\pi_i}{\sum_{K=1}^{n} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_K}{\sigma}\right)^2\right\}\pi_K}$$

(2)

$$\log\left(\frac{f_1(x)\pi_1}{f_2(x)\pi_2}\right) = \log\left(\frac{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right)\pi_1}{\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)\pi_2}\right)$$

$\sigma_1 = \sigma_2$ por Homogeneidad

$$\log\left(\frac{f_1(x)\pi_1}{f_2(x)\pi_2}\right) = \frac{(x-\mu_1)^2}{\sigma^2} - \frac{(x-\mu_2)^2}{\sigma^2} = 0$$

$$\frac{(x-\mu_1)^2}{\sigma^2} = \frac{(x-\mu_2)^2}{\sigma^2}$$

$$(x-\mu_1)^2 = (x-\mu_2)^2$$

$$(x-\mu_1)^2 - (x-\mu_2)^2 = 0$$

$$\cancel{x^2} - 2x\mu_1 + \mu_1^2 - \cancel{x^2} + 2x\mu_2 - \mu_2^2 = 0$$

$$2x(\mu_2 - \mu_1) = \mu_2^2 - \mu_1^2 = (\mu_2 - \mu_1)(\mu_2 + \mu_1)$$

$$x = \frac{\mu_2 + \mu_1}{2}$$

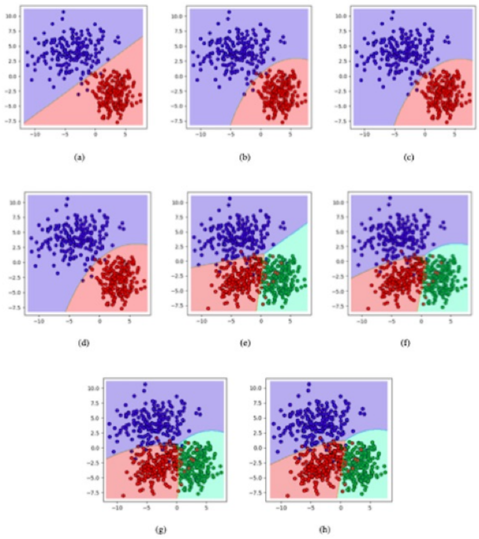
Ejercicio 2. Sean

$$\mu_1 = (-4, 4), \qquad \mu_2 = (3, -3), \qquad \mu_3 = (-3, 3)$$

y

$$\Sigma_1 = \begin{pmatrix} 10 & 1 \\ 1 & 5 \end{pmatrix} \qquad \Sigma_2 = \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix} \qquad \Sigma_3 = \begin{pmatrix} 6 & 1.5 \\ 1.5 & 4 \end{pmatrix}$$

¿Cuáles de las siguientes gráficas representan las fronteras de decisión?



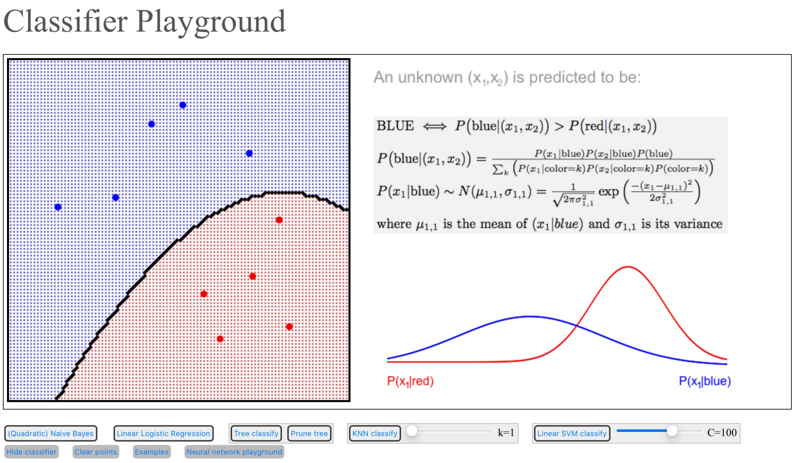
a) La gráfica (a) representa una frontera de decisión creada con LDA, es decir, clasifica los datos con un discriminante lineal y se tiene homogeneidad.

(b), (c), (d) Se tienen fronteras cuadráticas por QDA ya que no hay homogeneidad en los datos, es decir, sus matrices de varianza-covarianza son distintas.

(e), (f), (g), (h) Se tienen 3 conjuntos de datos. Por lo tanto, las fronteras de decisión entre cada conjunto de datos son lineales (LDA) si comparten la misma matriz de covarianza, de lo contrario, si no existe homogeneidad, la frontera de decisión es cuadrática asemejando a una parábola.

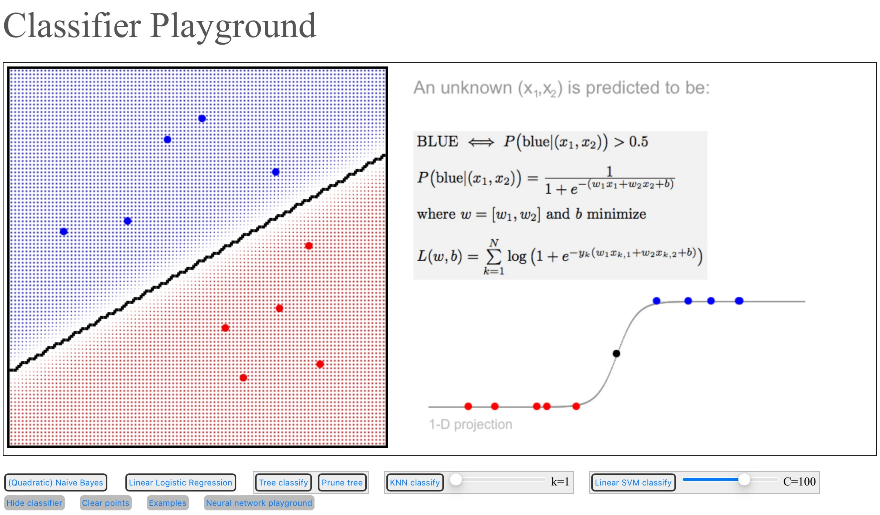
Ejercicio 3. Toma 5 puntos al azar de color rojo y 5 puntos al azar de color azul. Grafica las distintas gráficas de Voronoi con sus fronteras usando la siguiente applet: https://www.ccom.ucsd.edu/ cdeotte/programs/classify.html Explica cómo funciona la applet.

- Caso QDA.



La applet determina la posición de los puntos y utiliza la probabilidad Bayesiana a posteriori para determinar la probabilidad de clasificar cada punto en determinada región de color. Esto a través de asumir una distribución normal bivariada para los datos.

- Caso LDA. → Logit



En este caso la applet asigna la probabilidad de cada punto de pertenecer a un grupo de color con la forma de un modelo de regresión logístico, el cual origina una función sigmoide que clasifica los datos en sus extremos. Además, en este caso se supone homogeneidad.