

PRÁCTICO ANÁLISIS DISCRIMINANTE

1. a) Se desea discriminar entre dos poblaciones normales con vector de medias $(0, 0)$ y $(1, 1)$ y varianzas 2 y 4, con coeficiente de correlación lineal $\rho = 0,8$. Construir la función discriminante e interpretarla.
b) las probabilidades a priori son 0.7 para la primera población y 0.3 para la segunda. Calcular la función discriminante en cada caso.
2. a) Se desea discriminar entre 3 poblaciones normales con vectores de medias $(0, 0)$, $(1, 1)$ y $(0, 1)$ y varianzas 2 y 4 y coeficiente de correlación lineal $\rho = 0,5$. Calcular y dibujar las funciones discriminantes y hallar su punto de corte.
b) Si los costos de equivocarnos no son los mismos, de manera que el costo de clasificar en la tercera población cuando viene de la primera es el doble de los demás, calcular las funciones discriminantes. Asumimos que las probabilidades a priori son iguales
3. a) Si tenemos G grupos, denotamos la covarianza entre dos variables aleatorias x_j y $x_{j'}$ como $t(x_j, x_{j'}) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ij'} - \bar{x}_{j'})$, y $\bar{x}_{gj} = \frac{1}{n_g} \sum_{i=1}^{n_g} x_{ij}$ es la media de la variable j en el grupo g , probar que

$$t(x_j, x_{j'}) = \underbrace{\frac{1}{n} \sum_{g=1}^G \sum_{i=1}^{n_g} (x_{ij} - \bar{x}_{gj})(x_{ij'} - \bar{x}_{gj'})}_{\text{dentro de los grupos}=w(x_j, x_{j'})} + \underbrace{\sum_{g=1}^G \frac{n_g}{n} (\bar{x}_{gj} - \bar{x}_j)(\bar{x}_{gj'} - \bar{x}_{j'})}_{\text{entre grupos}=b(x_j, x_{j'})}$$

lo cual indica que la covarianza total es igual a la covarianza *dentro* de los grupos más la covarianza *entre* grupos.

- b) Con las notaciones usadas en teórico, probar que

$$T = W + B$$

4. Justifique los valores propios de $W^{-1}B$ son todos positivos. Para eso alcanza con ver que las matrices $W^{-1}B$ y $W^{-1/2}BW^{-1/2}$ tienen los mismos valores propios.
5. Pruebe que las variables z_j obtenidas son incorreladas. Ver Peña página 419.
6. Demostrar que es lo mismo obtener el mayor vector propio de $W^{-1}B$ y el menor de $T^{-1}W$
7. Buscar información sobre el test M de Box para testear si las matrices de varianzas-covarianzas son las mismas para las distintas subpoblaciones que se consideran.
8. Estudiar el apéndice 13.2 “Discriminación y Regresión” del libro de Peña.
9. En este ejercicio se trabaja sobre la base de datos **Smarket** de la librería ISLR
 - a) Buscar información sobre esta base de datos.
 - b) Realizar una regresión logística, un análisis discriminante lineal y un análisis discriminante cuadrático de la variable **Direction** sobre **Lag1** y **Lag2**, y comparar los resultados obtenidos.