

Parcial 2 de Multivariado

Punto 1(15 puntos)

En un agrupamiento de 26 unidades se trabajo con 2 variables designadas VarI y VarII y con el algoritmo k-means, obteniendo los resultados para 2, 3 y 4 grupos que se muestran en la siguiente tabla:

Dos grupos(Medias, suma de cuadrados dentro, tamaño)

Tres grupos(Medias, suma de cuadrados dentro, tamaño)

Cuatro grupos(Medias, suma de cuadrados dentro, tamaño)

Determinar el número óptimo de grupos a través del test F.

Punto 2(40 puntos)

Los clientes de un Banco fueron clasificados en cumplidores y fallidos. El gerente comercial realizó un análisis discriminante utilizando la información de 2 variables, designadas VarA y VarB, basado en una muestra de 70 clientes, de los cuales la mitad cumplió y la otra mitad no. En la tabla siguiente se presentan algunos de los resultados obtenidos:

Medias de grupos, Matriz de covarianzas combinada, coeficientes de la función de clasificación.

MediaGrupo1 = c(5,4.8); MediaGrupo2 = c(9, 3)

Sp = matrix(c(4.764,1.001,1.001,3.259),ncol=2,nrow=2,byrow=TRUE)

El gerente debe decidir si se concede el préstamo a dos nuevos solicitantes, quienes presentan los siguientes valores para las variables analizadas.

Solicitante1 = c(6,5); Solicitante2 = c(8,4)

- Escriba la expresión analítica de los coeficientes de la función discriminante lineal que minimiza la probabilidad total de clasificar mal y el vector de coeficientes estimados en este ejemplo.
- Para cada solicitante, obtenga la probabilidad a posteriori de pertenencia a cada grupo.
 - Explique cómo se obtiene la tasa de error cross-validada.

Punto 3(25 puntos)

A partir de la siguiente matriz de distancias (5x5)

- Obtenga la sucesión de agrupaciones que se alcanza con el método de Ligamiento promedio (UPGMA)
- Construya un diagrama de árbol jerárquico para las agrupaciones obtenidas.

Punto 4(20 puntos)

Para un conjunto de 26 observaciones y 6 variables se realizaron agrupaciones jerárquicas utilizando distintos métodos, representadas en los siguientes dendogramas.

Distancia euclidea, enlace(simple, completo, promedio,centroide) y ward

- ¿Con qué método se obtuvo el mejor agrupamiento de los datos?
- Describa brevemente el método seleccionado.
- Utilizando el dendograma ¿Cuántos grupos seleccionaría?¿Por qué?