

LABORATORIO 2 DE ANÁLISIS DISCRIMINANTE EN MATLAB

Este ejercicio consiste es para aplicar método análisis discriminante en MATLAB para más de dos grupos. En este caso la prueba será con cuatro variables previamente clasificadas y se corresponde con el incluido en las notas disponibles en este sitio.

La tabla se corresponde con cuatro agrupaciones hipotéticas que muestra dos variables x , x_1 y x_2 .

	Grupo 1		Grupo 2		Grupo 3		Grupo 4	
Obs.	x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2
1	1	3	11	3	1	13	11	13
2	2	1	12	1	2	11	12	11
3	4	1	14	1	4	11	14	11
4	5	3	15	3	5	13	15	13
5	4	4,5	14	4,5	4	14,5	14	14,5
6	2	4,5	12	4,5	2	14,5	12	14,5
7	3	1,5	13	1,5	3	11,5	13	11,5
8	4,5	1,5	14,5	1,5	4,5	11,5	14,5	11,5
9	4,5	3	14,5	3	4,5	13	14,5	13
10	3	4	13	4	3	14	13	14
11	2	3,5	12	3,5	2	13,5	12	13,5
12	2	2	12	2	2	12	12	12
13	3	3	13	3	3	13	13	13

Igual que en el laboratorio anterior, el ejercicio consiste en crear un procedimiento en MATLAB que permita clasificar satisfactoriamente a un grupo de observaciones nuevas de acuerdo a una función discriminante.

El procedimiento emplea la función:

[clases,error] =classify(P,E,GE)

donde

clases: es el resultado de clasificar los valores nuevos dispuestos en la matriz prueba

error: nivel de error en clasificación

P: es la matriz previamente preparada de observaciones para probar el modelo

E: matriz de entrenamiento formada por observaciones originales de x_1 y x_2

GE: vector formado por las clases de las observaciones en E.

GP: vector formado por las clases de las observaciones en P.

Inicialmente, desde los datos suministrados en Excel (en lab_2_subhash), *organice una tabla con tres columnas de 58 observaciones formada por los valores de x_1 , x_2 y el grupo asociado a cada par x* , para luego transformarlo a un archivo tipo texto separado por tabuladores (acordarse del cambio de las comas a punto en ese archivo para los valores con decimales). Se carga el archivo de datos que se corresponde con la tabla anterior con el nombre que haya indicado.

load 'nombre.txt'

datos=nombre;

En **datos** deben existir 52 observaciones y 3 columnas

size(datos)

Organizamos las observaciones de x_1 y x_2 en una matriz XY de 52 filas y 2 columnas

XY=[datos(:,1);datos(:,2)];

Luego el vector de 52 elementos correspondientes con el grupo asignado a cada observación

G=[datos(:,3)];

Definimos una matriz de observaciones para modelo **classify**

E=[XY(1:8,:);XY(14:21,:);XY(27:34,:);XY(40:47,)];

y la matriz de prueba con observaciones de prueba para crear la matriz de confusión

P=[XY(9:13,:);XY(22:26,:);XY(35:39,:);XY(48:52,)];

El vector de grupos asociados a cada para en E

GE=[G(1:8,:);G(14:21,:);G(27:34,:);G(40:47,)];

Y el correspondiente a los grupos de la matriz de prueba

GP=[G(9:13,:);G(22:26,:);G(35:39,:);G(48:52,)];

Aplique los siguientes comandos para verificar las dimensiones de cada arreglo y la organización de los datos.

whos

E'

P'

GE'

GP'

Ahora aplique el comando para realizar el análisis discriminante multiple

[clases,error]=classify(P,E,GE)

Observe los resultados en **clases**

Ahora veamos algunas representaciones

gscatter(E(:,1),E(:,2),GE)

Gráfico con las observaciones de prueba y los grupos a los cuales fueron clasificadas las observaciones mediante el modelo.

gscatter(P(:,1),P(:,2),clases,'gm','o*')

hold

Current plot held

Para comparar con el gráfico anterior se usa la misma función con las observaciones de prueba y los grupos a los cuales realmente observados. Tenemos la siguiente superposición de gráficas

gscatter(P(:,1),P(:,2),GP,'br','.x')

Para crear la matriz de confusión, y observar que tan buena fue la clasificación. De acuerdo a lo estimado y a lo observado, el mismo conjunto de observaciones se agrupan en el arreglo **balance**, donde la primera columna son los valores observados de los grupos y en la segunda, los valores estimados

balance=[GP(:) clases(:)];

Las frecuencias para cada vector de grupos es

tabulate(balance(:,1))

tabulate(balance(:,2))

La matriz de confusión resultante es

confusion=crosstab(balance(:,1),balance(:,2))

Determinamos los aciertos y errores en la clasificación en la matriz confusión que debería ser de orden 4x4.