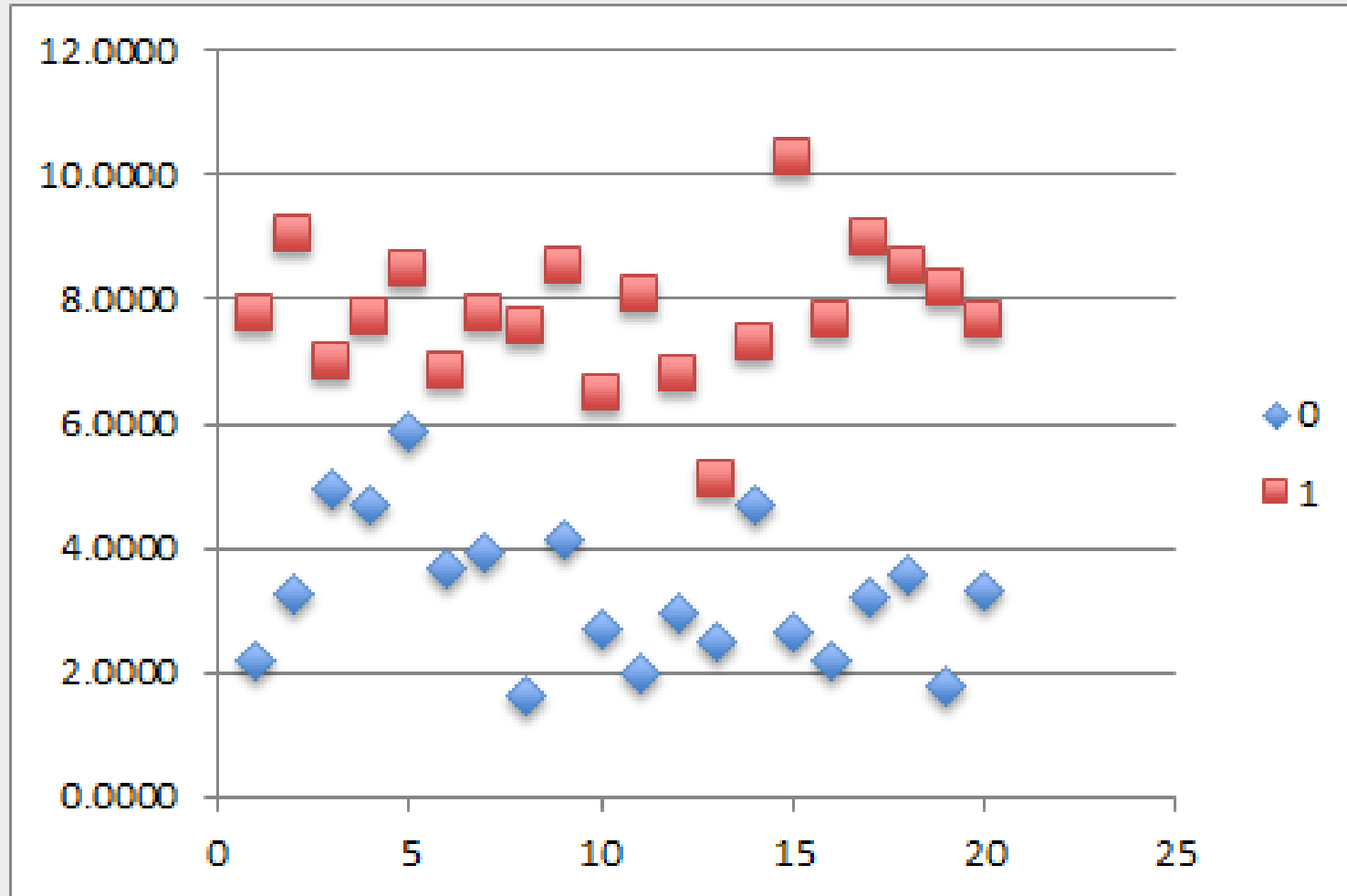


TA 5- Ej1 Análisis Discriminante Lineal (1)

- recordar que el LDA **asume** que:
 - las variables de entrada tienen distribución normal.
 - la varianza de cada variable de entrada, agrupada por cada clase de la salida, es la misma
- **utilizando una planilla electrónica,**
- utilizaremos un conjunto de datos con 1 atributo, y 1 salida con dos clases
- insertar los datos contenidos en el archivo “TA5-DATASET.CSV”
 - la variable Y tiene dos Clases: 0 y 1
 - los valores de la variable X han sido generados con distribuciones normales y varianzas similares

TA 5- Ej1 Análisis Discriminante Lineal (2)

- grafica los datos, separando en las dos clases de la salida



TA5 - Ej 1 - Análisis Discriminante Lineal (3)

- Calcula $P(Y=0)$ y $P(Y=1)$ (conteo)
- para $Y=0$,
 - calcula la media de los X y, para cada X , $(x - \text{media}(x))^2$
 - totaliza las diferencias cuadráticas
- repite para $Y=1$

Probabilidad $P(y=0)$	$P(y=1)$		
0.5	0.5		
X, Y=0	media(x)	$(x - \text{media}(x))^2$	Suma
2.1936	3.2990	1.2219	
3.2731		0.0007	
X, Y=0	media(x)	$(x - \text{media}(x))^2$	Suma
2.1936	3.2990	1.2219	
3.2731		0.0007	
1.7922		2.2704	
3.3280		0.0008	25.3357
X, Y=1	media(x)	$(x - \text{media}(x))^2$	Suma
7.7691	7.7871	0.0003	
9.0430		1.5773	
7.6416		0.0212	22.5028

TA5 - Ej 1 - Análisis Discriminante Lineal (4)

- Calcula la varianza con los datos de diferencias hallados (aquí $K = 2$ y $n = 40$):

$$\sigma^2 = \frac{1}{n - K} \times \sum_{i=1}^n (x_i - \text{mean}_k)^2$$

- ya tienes todos los parámetros del modelo!

$$D_k(x) = x \times \frac{\text{mean}_k}{\sigma^2} - \frac{\text{mean}_k^2}{2 \times \sigma^2} + \ln(P(k))$$

TA5 - Ej 1 - Análisis Discriminante Lineal (5)

- Predicción y error
 - utilizar los mismos datos de entrada de X
 - en las 2 siguientes columnas calcular los discriminantes para Y=0 e Y=1

$$D_k(x) = x \times \frac{mean_k}{sigma^2} - \frac{mean_k^2}{2 \times sigma^2} + \ln(P(k))$$

- calcular la predicción de clase (comparar los discriminantes)
- agregar una columna con las Y originales, calcular el error de predicción en cada ejemplo, y al final, la **exactitud** de la predicción total