



PA
Análisis de Datos

Prof: Max Chacón
Marzo 2016

1. Para caracterizar billetes falsos los bancos suizos realizaron un análisis que consistía en tomar medidas de los billetes. Para el análisis tenía tres grupos diferente de billetes. Originales de papel, originales de plástico y billetes falsos. Cada billete fue caracterizado por las siguientes variables:

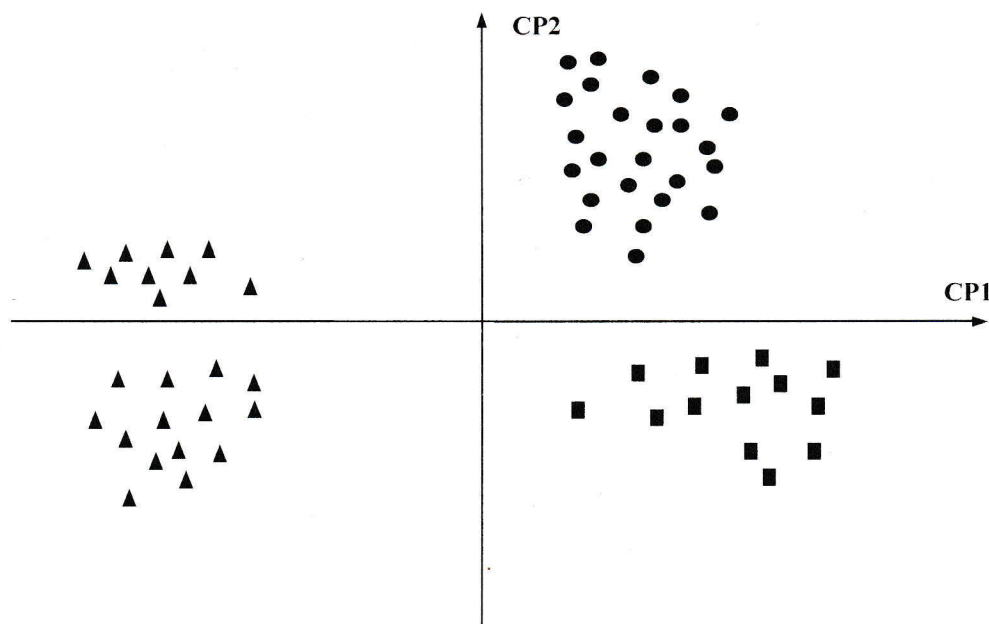
Lon: Longitud del billete
LD: Largo de la Diagonal del billete
AI: Ancho Izquierdo del billete
AD: Ancho Derecho del billete
AMI: Ancho Margen Inferior del billete
AMS: Ancho Margen Superior del billete.

A continuación, se realizó un análisis de componentes principales con los siguientes resultados: - Valores propios: 2,58; 1,34; 0,76; 0,56; 0,50; 0,26.

- Vectores propios para las dos primeras componentes.

	CP1	CP2
Lon	0,395	0,799
LD	0,207	0,345
AI	0,445	-0,263
AD	0,411	-0,375
AMI	0,347	-0,072
AMS	0,560	-0,163

La distribución de los diferentes tipos de billetes se muestran en la figura (originales de papel en círculos, originales de plástico en cuadrados y falsos en triángulos):



Se realizó un trabajo similar con monedas midiendo 6 variables de tamaño de éstas y los valores propios del análisis fueron los siguientes: 1,96; 1,54; 1,09; 0,73; 0,40; 0,28.

- Determine el porcentaje de validez del análisis.
- Interprete cada una de las componentes
- Identifique las principales características de los billetes originales.
- Determine si existen diferencias entre las falsificaciones.
- Para el análisis de las monedas, ¿se logrará tener una precisión similar a la de los billetes?. (1,5)

2. En una gran variedad de problemas de bioinformática, se requiere determinar grupos de genes que intervienen en una determinada enfermedad. Los genes están formados por un alfabeto básico que contiene 4 letras (bases) A, G, C, T. Muchas veces con un trozo de gen (aproximadamente 8 bases "letras") es posible caracterizar un gen. Para la siguiente tabla determine: a) Un método para medir distancia entre genes. b) Construya una matriz de distancia entre los genes. c) Determine el dendrograma

(1,5)

Gen 1	G	A	T	A	C	A	T	T
Gen 2	G	A	T	A	C	A	T	A
Gen 3	G	A	T	A	C	T	A	C
Gen 4	C	T	A	A	G	G	G	G
Gen 5	C	T	C	A	G	G	G	G
Gen 6	G	A	T	T	T	C	C	G
Gen 7	G	A	T	T	A	C	C	G

3. Se requiere determinar entre las características: Depresión de segmento ST e Inclinación del segmento ST, la que clasifica mejor la patología. Se cuenta con 1.000 casos para la Depresión del segmento y 800 para el caso de Inclinación, como se muestra en las siguientes tablas.

	Con Isquemia	Sin Isquemia
Sin depresión ST	200	400
Con depresión ST	200	200

	Con Isquemia	Sin Isquemia
Sin Inclinación ST	48	552
Con Inclinación ST	200	0

(1,5)

4. a) La respuesta al escalón de un sistema se puede expresar como $y(n) = u(n) * h(n)$.

Expresé $y(n)$ en términos de $h(n)$ exclusivamente. Como queda esta representación, si el sistema es causal. Calcule además, $y(n)$ cuando $h(n) = u(n) - u(n-3)$.

b) Qué relación existe entre la convolución de dos señales $y(n) * x(n)$ y su correlación r_{xy} . (1,5).