

## A PRIORI

| ID | TRANSACTIONS | SOPORT MINIMO = 3 |  |   |
|----|--------------|-------------------|--|---|
| 1  | 1, 2, 3      | $k=1$             | $k=2$  | $k=3$                                       |
| 2  | 2, 3, 4      | $1 \Rightarrow 5$ | <del>1,2 <math>\Rightarrow</math> 3</del>                                      | <del>1,3,4 <math>\Rightarrow</math> 3</del> |
| 3  | 1, 2, 4      | $2 \Rightarrow 6$ | <del>1,3 <math>\Rightarrow</math> 4</del>                                      | <del>2,3,4 <math>\Rightarrow</math> 5</del> |
| 4  | 2, 3         | $3 \Rightarrow 7$ | <del>1,4 <math>\Rightarrow</math> 4</del>                                      | <del>4,2,3 <math>\Rightarrow</math> 2</del> |
| 5  | 1, 3, 4      | $4 \Rightarrow 6$ | <del>2,3 <math>\Rightarrow</math> 5</del>                                      | <del>2,4 <math>\Rightarrow</math> 4</del>   |
| 6  | 2, 3, 4      |                   | <del>3,4 <math>\Rightarrow</math> 5</del>                                      |   |
| 7  | 1, 3, 4      |                   |  |   |
| 8  | 1, 2, 3, 4   |                   | $\left\{ \begin{array}{l} 1,3 \\ 1,4 \\ 2,3 \\ 2,4 \\ 3,4 \end{array} \right.$ |   |

## ALGORITMO K-MEANS

$$P_1 = (2, 3)$$

$$P_2 = (3, 5) \quad C_1 = (4, 3)$$

$$P_3 = (5, 8) \quad C_2 = (9, 7)$$

$$P_4 = (7, 3)$$

$$P_5 = (6, 5)$$

$$P_6 = (11, 1)$$

CON C1:

$$D(P_1, C_1) = \sqrt{(2-4)^2 + (3-3)^2} = 2$$

$$D(P_2, C_1) = \sqrt{(3-4)^2 + (5-3)^2} = 2'23$$

$$D(P_3, C_1) = \sqrt{(5-4)^2 + (8-3)^2} = 5'09$$

$$D(P_4, C_1) = \sqrt{(7-4)^2 + (3-3)^2} = 3$$

$$D(P_5, C_1) = \sqrt{(6-4)^2 + (5-3)^2} = 6'32$$

$$D(P_6, C_1) = \sqrt{(11-4)^2 + (1-3)^2} = 8'06$$

CON C2:

$$D(P_1, C_2) = \sqrt{(2-8)^2 + (3-7)^2} = 7'21$$

$$D(P_2, C_2) = \sqrt{(3-7)^2 + (5-7)^2} = 5'38$$

$$D(P_3, C_2) = \sqrt{(5-7)^2 + (8-7)^2} = 3'16$$

$$D(P_4, C_2) = \sqrt{(7-7)^2 + (3-7)^2} = 4'12$$

$$D(P_5, C_2) = \sqrt{(10-7)^2 + (5-7)^2} = 2'82$$

$$D(P_6, C_2) = \sqrt{(11-7)^2 + (1-7)^2} = 3$$

MATRIZ DISTANCIAS

$$\begin{pmatrix} 0 & 2'23 & 5'09 & 3 & 6'32 & 8'06 \\ 2'21 & 0 & 5'38 & 3'16 & 4'12 & 2'82 \\ 5'38 & 2'21 & 0 & 4'12 & 2'82 & 3 \end{pmatrix}$$

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \Rightarrow \begin{array}{l} C_1 = (4, 3'66) \\ C_2 = (8'66, 6, 66) \end{array}$$

REPETIMOS

$$D(P_1, C_1) = \sqrt{(2-4)^2 + (3-3.66)^2} = 2'10$$

$$D(P_2, C_1) = \sqrt{(3-4)^2 + (5-3.66)^2} = 1'67$$

$$D(P_3, C_1) = 4'45$$

$$D(P_4, C_1) = 3'07$$

$$D(P_5, C_1) = 6'14$$

$$D(P_6, C_1) = 7'75$$

$$D(P_1, C_2) = \sqrt{(2-8.66)^2 + (3-6.66)^2} = 9'59$$

$$D(P_2, C_2) = 5'89$$

$$D(P_3, C_2) = 3'89$$

$$D(P_4, C_2) = 4'01$$

$$D(P_5, C_2) = 2'13$$

$$D(P_6, C_2) = 2'36$$

## MATRIZ DE DISTANCIAS

$$\begin{pmatrix} 2'10 & 1'67 & 4'45 & 3'07 & 6'14 & 7'75 \\ 7'59 & 5'89 & 3'89 & 4'01 & 2'13 & 2'36 \end{pmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$C_1 = (4.25, 4.75)$$

$$C_2 = (10, 5, 6)$$

REPETIMOS

$$D(P_1, C_1) = \sqrt{(2-4.25)^2 + (3-4.75)^2} = 2'81$$

$$D(P_2, C_1) = 1'27$$

$$D(P_3, C_1) = 3'33$$

$$D(P_4, C_1) = 3'25$$

$$D(P_5, C_1) = 5'75$$

$$D(P_6, C_1) = 7'11$$

$$D(P_1, C_2) = \sqrt{(2-10.5)^2 + (3-6)^2} = 9'01$$

$$D(P_2, C_2) = 7'56$$

$$D(P_3, C_2) = 5'85$$

$$D(P_4, C_2) = 4'60$$

$$D(P_5, C_2) = 1'11$$

$$D(P_6, C_2) = 4'11$$

## MATRIZ DE DISTANCIAS

$$\begin{pmatrix} 2'81 & 1'27 & 3'33 & 3'25 & 5'75 & 7'11 \\ 9'01 & 7'56 & 5'85 & 4'60 & 1'11 & 4'11 \end{pmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \rightarrow \text{no cambia, hemos encontrado los centroides.}$$

CLUSTERIN TEKARALICO

→  
MAS LERAMO →

|   | A    | B    | C     | D    | E    | F    | G |
|---|------|------|-------|------|------|------|---|
| A | 0    |      |       |      |      |      |   |
| B | 2,15 | 0    |       |      |      |      |   |
| C | 0,78 | 1,53 | 0     |      |      |      |   |
| D | 1,07 | 1,14 | 0,43  | 0    |      |      |   |
| E | 0,85 | 1,38 | 0,21  | 0,29 | 0    |      |   |
| F | 1,16 | 1,01 | 0,55  | 0,22 | 0,41 | 0    |   |
| G | 1,56 | 2,83 | 1,856 | 2,04 | 2,02 | 2,05 | 0 |

|     | A      | B      | CE      | D      | F      | G      |
|-----|--------|--------|---------|--------|--------|--------|
| A   | 0      |        |         |        |        |        |
| B   | ✓ 2,15 | 0      |         |        |        |        |
| CE  | ✓ 0,78 | ✓ 1,38 | 0       |        |        |        |
| → D | ✓ 1,07 | ✓ 1,14 | ✓ 0,29  | 0      |        |        |
| → F | ✓ 1,16 | ✓ 1,01 | ✓ 0,41  | ✓ 0,22 | 0      |        |
| G   | ✓ 1,56 | ✓ 2,83 | ✓ 1,856 | ✓ 2,04 | ✓ 2,02 | ✓ 2,05 |

|      | A      | B      | CE      | DF     | G |
|------|--------|--------|---------|--------|---|
| A    | 0      |        |         |        |   |
| B    | ✓ 2,15 | 0      |         |        |   |
| → CE | ✓ 0,78 | ✓ 1,38 | 0       |        |   |
| → DF | ✓ 1,07 | ✓ 1,01 | ✓ 0,29  | 0      |   |
| G    | ✓ 1,56 | ✓ 2,83 | ✓ 1,856 | ✓ 2,04 | 0 |

|            | A      | B      | (CE)(DF) | G      |   |
|------------|--------|--------|----------|--------|---|
| A          | 0      |        |          |        |   |
| B          | ✓ 2,15 | 0      |          |        |   |
| (CE), (DF) | ✓ 0,78 | ✓ 1,01 | ✓ 0      |        |   |
| G          | ✓ 1,56 | ✓ 2,83 | ✓ 1,856  | ✓ 2,04 | 0 |

PCA

- Nedan tabla con componentes y autorrelaciones.
- Sumamos todos los autorrelaciones.
- Dividimos cada autorrelacion entre la suma total.
- Vamos multiplicando el % y juntando los componentes.

(AT PRIORI)

Soporte mínimo: 0'5

Confianza: 0'9

| ID | Comercios visitados                       |
|----|---|
| 1  | Zapatería, Electrónica, Alimentación      |
| 2  | Electrónica, Electrónica, Ropa            |
| 3  | Zapatería, Electrónica, Electrónica, Ropa |
| 4  | Electrónica, Ropa.                        |

PASO 1: CONTAR CADA ELEMENTO

| Elemento     | Soporte  |
|--------------|--|
| Zapatería    | $2 \Rightarrow 2/5 = 0'4$ (menor que soporte = se eliminó) |
| Electrónica  | $3 \Rightarrow 3/5 = 0'6 >$ soporte                        |
| Alimentación | $1 \Rightarrow 1/5 = 0'2 <$ soporte                        |
| Ropa         | $3 \Rightarrow 3/5 = 0'6 >$                                |
| Electrónica  | $3 \Rightarrow 3/5 = 0'6 >$                                |

PASO 2: JUNTANDO LOS RESIDUOS

PASO 3: EVALUAMOS REGLAS

$$R \rightarrow E \Rightarrow 0'6/0'6 = 100\%$$

$$E \rightarrow R \Rightarrow 0'6/0'6 = 100\%$$

| Elemento                 | Soporte                     |
|--------------------------|-----------------------------|
| Electrónica, Ropa        | $2 \Rightarrow 2/5 = 0'4 <$ |
| Electrónica, Electrónica | $2 \Rightarrow 2/5 = 0'4 <$ |
| Ropa, Electrónica        | $3 \Rightarrow 3/5 = 0'6 >$ |

## K-Nearest Neighbors

Dado CONJUNTO DE ENTRENAMIENTO

|       | CARGO<br>SECO | ANCHO<br>AWL | PELVO<br>P | ALLO<br>A | CLASE      |
|-------|---------------|--------------|------------|-----------|------------|
| $x_1$ | 5'1           | 3'5          | 1'4        | 0'2       | setosa     |
| $x_2$ | 4'0           | 3'0          | 1'4        | 0'2       | setosa     |
| $x_3$ | 4'7           | 3'0          | 1'3        | 0'2       | setosa     |
| $x_4$ | 7'0           | 3'2          | 4'7        | 1'4       | versicolor |
| $x_5$ | 3'4           | 3'2          | 4'5        | 1'5       | v          |
| $x_6$ | 6'9           | 3'1          | 4'9        | 1'5       | -v         |

donde nros individuos  $\Rightarrow (5'0, 3'3, 1'4, 0'2) = x$

$$D(x_1, x) = \sqrt{(5'1 - 5'0)^2 + (3'5 - 3'3)^2 + (1'4 - 1'4)^2 + (0'2 - 0'2)^2} = 0'22$$

$$D(x_2, x) = \sqrt{(4'0 - 5'0)^2 + (3'0 - 3'3)^2 + (1'4 - 1'4)^2 + (0'2 - 0'2)^2} = 1'04$$

$$D(x_3, x) = \sqrt{(4'7 - 5'0)^2 + (3'0 - 3'3)^2 + (1'3 - 1'4)^2 + (0'2 - 0'2)^2} = 0'43$$

$$D(x_4, x) = 4'04$$

$$D(x_5, x) = 3'72$$

$$D(x_6, x) = 4'19$$

### CLASIFICACIONES

$k=1 \Rightarrow$  el proximo mas a  $x_i \sim$  clase  $x =$  setosa

$k=2 \Rightarrow x_1 \wedge x_2 \Rightarrow$  los dos son setosa  $\Rightarrow$  clase  $x =$  setosa

$k=3 = x_1, x_2 \wedge x_3 \Rightarrow$  los 3 son setoso  $\Rightarrow$  clase  $x =$  setoso.

ALGORITMO  
K-MEANS

$k=2$

|        |   | (1,3) (3,7) |
|--------|---|-------------|
|        |   | C1 C2       |
| ALTURA |   | ALTURA      |
| P1     | 1 | 1           |
| P2     | 3 | 2           |
| P3     | 4 | 4           |
| P4     | 6 | 4           |

DISTANCIA PUNTO A CENTROIDE 1

$$D(P_1, C_1) = \sqrt{(1-1)^2 + (1-3)^2} = 2$$

$$D(P_2, C_1) = \sqrt{(3-1)^2 + (2-3)^2} = 2\sqrt{2}$$

$$D(P_3, C_1) = \sqrt{(4-1)^2 + (4-3)^2} = 3\sqrt{2}$$

$$D(P_4, C_1) = \sqrt{(6-1)^2 + (4-3)^2} = 5\sqrt{2}$$

DISTANCIA PUNTO A CENTROIDE 2

$$D(P_1, C_2) = \sqrt{(1-3)^2 + (1-7)^2} = 6\sqrt{2}$$

$$D(P_2, C_2) = \sqrt{(3-3)^2 + (2-7)^2} = 5$$

$$D(P_3, C_2) = \sqrt{(4-3)^2 + (4-7)^2} = 3\sqrt{2}$$

$$D(P_4, C_2) = \sqrt{(6-3)^2 + (4-7)^2} = 3\sqrt{10}$$

DONEMOS EN MATRIZ

$$\begin{matrix} & P_1 & P_2 & P_3 & P_4 \\ C_1 & \begin{pmatrix} 2 & 2\sqrt{2} & 3\sqrt{2} & 5\sqrt{2} \\ 6\sqrt{2} & 5 & 3\sqrt{2} & 3\sqrt{2} \end{pmatrix} & \xrightarrow{\substack{\text{MATRIZ} \\ \text{COLUMNS}}} & \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ C_2 & & & \xrightarrow{\substack{\text{Por columna} \\ \text{el mas bajo}}} \end{matrix}$$

C1 → damos los valores X del punto con 1, los numero y dividimos entre el nº de vueltas.  
 - lo mismo con Y.

C2 → lo mismo

$$C_1 = \left( \frac{(1+3+4)}{3}, \frac{(1+2+4)}{3} \right) = C_1 = (2,66,2\sqrt{2})$$

$$C_2 = \left( \frac{6}{4}, \frac{4}{4} \right) = C_2 = (6,4)$$

} REPETIR LOS MISMO CONDICIONES.

PARAR LOS CUANDO NO

HAYA NUEVOS  
CENTROIDES

REGRESIÓN LINEAR  
SIMPLÍCIA

- Calcular  $X_i, Y_i, X_i^2, Y_i^2, X_i \cdot Y_i$

- calcular  $\bar{X} \Rightarrow$  Sumatorio de  $X_i / n^{\circ}$  individuos (media)

calcular  $\bar{Y} \Rightarrow$  val con  $Y_i$ .

$$\text{calculo } Sx^2 = \frac{\sum X_i^2}{n^{\circ}} - \bar{X}^2$$

$$\text{calculo } Sy^2 = \frac{\sum Y_i^2}{n^{\circ}} - \bar{Y}^2$$

$$\text{calculo } SXY = \frac{\sum X_i Y_i}{n^{\circ}} - \bar{X} \cdot \bar{Y}$$

$$-\text{calcular } r = \frac{SXY}{\sqrt{Sx^2 \cdot Sy^2}} \neq \text{entr } 0 \text{ o } 1.$$

- Modelo de Regresión

$$X - \bar{X} = \frac{SXY}{Sy^2} \cdot (Y - \bar{Y}) \Rightarrow \text{despejado } X$$

$$Y - \bar{Y} = \frac{SXY}{Sx^2} \cdot (X - \bar{X}) \Rightarrow \text{despejado } Y$$

EJERCICIO EN EXCEL

| REGRESION LINEAL SIMPLE |               |      |      |     |     |
|-------------------------|---------------|------|------|-----|-----|
| Nº Clientes (X)         | Distancia (Y) | X^2  | Y^2  | X*Y |     |
| 15                      | 8             | 225  | 64   | 120 |     |
| 19                      | 7             | 361  | 49   | 133 |     |
| 25                      | 6             | 625  | 36   | 150 |     |
| 23                      | 4             | 529  | 16   | 92  |     |
| 34                      | 2             | 1156 | 4    | 68  |     |
| SUMATORIOS              | 116           | 27   | 2896 | 169 | 563 |

|         |              |  |  |  |  |
|---------|--------------|--|--|--|--|
| MEDIA X | 23,2         |  |  |  |  |
| MEDIA Y | 5,4          |  |  |  |  |
| SX^2    | 40,96        | $\sum X_i^2/N - \bar{X}^2$                       |  |  |  |
| SY^2    | 4,64         | LO Mismo que SX^2 pero con Y                     |  |  |  |
| SXY     | -12,68       | $\sum X_i Y_i / N - \bar{X} \cdot \bar{Y}$       |  |  |  |
| R       | -0,919772222 | Muy relacionados inversamente proporcional       |  |  |  |
| R^2     | 0,84598094   | $\Rightarrow \frac{SXY}{\sqrt{SX^2 \cdot SY^2}}$ |  |  |  |

MODELOS

$$x - \bar{x} = \frac{SXY}{SY^2} \cdot (y - \bar{y}) \Rightarrow x = \frac{-12,68}{4,64} \cdot (y - 5,4) + 23,2 \Rightarrow x = -2,732 \cdot y + 37,95$$

$$y - \bar{y} = \frac{SXY}{SX^2} \cdot (x - \bar{x}) \Rightarrow y = \frac{-12,68}{40,96} \cdot (x - 23,2) + 5,4 \Rightarrow y = -0,31x + 12,58$$

PRESTAMO - } SI, NO

**CLASIFICACIÓN P-ADADA  
EN BAYES**

| SEXO   | SI  | NO  |
|--------|-----|-----|
| HOMBRE | 1/3 | 3/5 |
| MUJER  | 2/3 | 2/5 |

| RENTA   | SI  | NO   |
|---------|-----|------|
| ALTA    | 1/3 | 1/25 |
| BAJA    | 1/3 | 2/5  |
| MEDIANA | 1/3 | 2/5  |

| EDAD    | SI  | NO  |
|---------|-----|-----|
| JUVENIL | 0/3 | 2/5 |
| MEDIA   | 2/3 | 2/5 |
| MAYOR   | 1/3 | 1/5 |

$\epsilon = \{ \text{Hombre, Baja, Juvenil} \}$

$$P(\text{NO}|\epsilon) = P(\text{Hombre}|\text{No}) \cdot P(\text{Baja}|\text{No}) \cdot P(\text{Juvenil}|\text{No}) \cdot P(\text{No})$$

$$P(\text{NO}|\epsilon) = 3/5 \cdot 2/5 \cdot 2/5 \cdot 5/8 = 0.06$$

$$P(\text{SI}|\epsilon) = P(\text{Hombre}|\text{SI}) \cdot P(\text{B}|\text{SI}) \cdot P(\text{J}|\text{SI}) \cdot P(\text{S})$$

$$P(\text{SI}|\epsilon) = 1/3 \cdot 1/3 \cdot 0/3 \cdot 3/8 = 0$$

$$\sum P = 0.06 + 0 = 0.06$$

$$\text{Si } 0.06 = \rightarrow \boxed{P(\text{Hombre}|\text{No}) = 100\%}$$

EJERCICIO BAYES

| COLOR    | MANZANA | NARANJA |
|----------|---------|---------|
| ROJO     | 2/3     | 0/3     |
| AMARILLO | 1/3     | 1/3     |
| NARANJA  | 0/3     | 2/3     |

| TEXTURA | MANZANA | NARANJA |
|---------|---------|---------|
| RUGOSA  | 1/3     | 2/3     |
| SIAMB   | 2/3     | 1/3     |

$E = \{R\}$  Rojo, Rugoso

$$P(N|E) = P(R|N) \cdot P(R|N) \cdot P(N) = 0 \cdot 2/3 \cdot 3/6 = 0$$

$$P(M|E) = P(R|M) \cdot P(R|M) \cdot P(M) = 2/3 \cdot 1/3 \cdot 3/6 = 0'11$$

$$\sum P = 0'11 + 0 = 0'11$$

$$\text{Si } 0'11 = 1 \rightarrow [P(\text{Montaña}|E) = 100\%]$$

$$x = 0'11 - 1$$

$$= \frac{0'11}{0'11} = 1$$

$$y = 0$$

$E = \{A\}$  Amarillo, suave

$$P(N|E) = P(A|N) \cdot P(S|N) \cdot P(N) = 1/3 \cdot 1/3 \cdot 3/6 = 0'055$$

$$P(M|E) = P(A|M) \cdot P(S|M) \cdot P(M) = 1/3 \cdot 2/3 \cdot 3/6 = 0'111$$

$$\sum P = 0'111 + 0'055 = 0'166$$

$$\text{Si } 0'166 = 1 \rightarrow x = \frac{0'111}{0'166} = 0'6894 \Rightarrow [P(M|E) = 68'94\%] + \underline{\text{señal manzana}}$$

$$y = 1 - x = 1 - 0'6894 = 0'3105 \Rightarrow P(N|E) = 0'3105\%$$

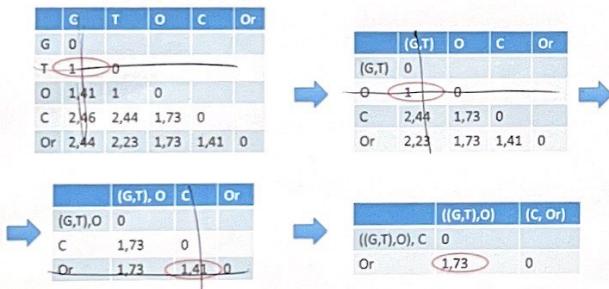
## Algoritmos: Clústering jerárquico

Ejemplo:

| Animal       | Orejas  | Movilidad  | Hábitat   | Pico |
|--------------|---------|------------|-----------|------|
| Gato         | Punta   | Cuadrúpedo | Doméstico | No   |
| Tigre        | Punta   | Cuadrúpedo | Salvaje   | No   |
| Oso          | Redonda | Cuadrúpedo | Salvaje   | No   |
| Cuervo       | No      | Bípedo     | Salvaje   | Si   |
| Ornítorrinco | No      | Cuadrúpedo | Salvaje   | Si   |

DIVISIÓN  
EXCLUIDA →  
 $D(\text{columna}, \text{fila})$

## Algoritmos: Clústering jerárquico



## Algoritmos: Clústering jerárquico

Ejemplo:

| Animal       | Orejas | Movilidad | Hábitat | Pico |
|--------------|--------|-----------|---------|------|
| Gato         | 1      | 1         | 1       | 1    |
| Tigre        | 1      | 1         | 2       | 1    |
| Oso          | 2      | 1         | 2       | 1    |
| Cuervo       | 3      | 2         | 2       | 2    |
| Ornítorrinco | 3      | 1         | 2       | 2    |

|    | G    | T    | O    | C    | Or |
|----|------|------|------|------|----|
| G  | 0    |      |      |      |    |
| T  | 1    | 0    |      |      |    |
| O  | 1,41 | 1    | 0    |      |    |
| C  | 2,46 | 2,44 | 1,73 | 0    |    |
| Or | 2,44 | 2,23 | 1,73 | 1,41 | 0  |

## Algoritmos: Clústering jerárquico

Interpretación:  
 $((G,T),O), (C,Or)$

El gato, el tigre se parecen mucho (uno es doméstico y otro salvaje) y el oso se parece a ellos en parte. Luego jerárquicamente encontramos al cuervo y al ornitorrinco (tienen pico) pero no se parecen en prácticamente nada al resto.