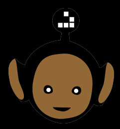
Notas Tema 2 AED

Rodrigo Castillo

5 de septiembre de 2020



1. Distancia Euclideana

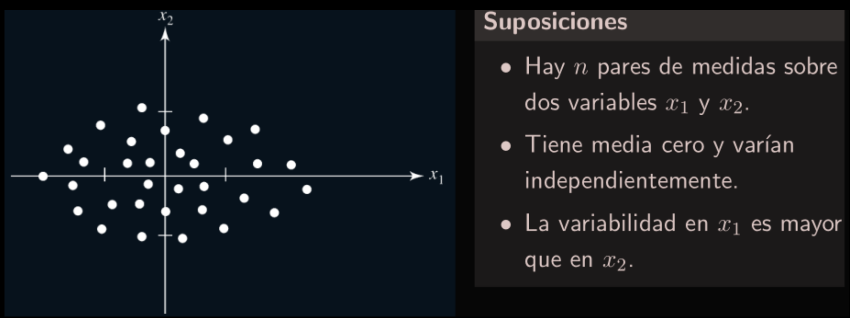
- la mayoria de las tecnicas multivariables dependen del concepto de distancia

sean $P = (x_1, x_2, x_3 ..., x_p)$ y $Q = (y_1, y_2, y_3, ..., y_p)$ se tiene que la distancia entre P y Q está dada como :

$$d(P,Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_p - y_p)^2} \tag{1}$$

- la distancia euclideana sirve para medir distancias entre 2 puntos pero en estadística no sirve, se usa una distancia llamada Distancia Estadística
- no funciona porque cada coordenada pesa lo mismo
- en la estadistica es bueno darle mayor peso a las variables que varian menos y menor peso a aquellas que varian mas
- la distancia estadístca tiene en cuenta tanto el peso de la variacion como el peso de las variables

2. Distancia Estadistica



- los puntos respecto al origen en la direccion de x_1 son mas comunes que los puntos a la misma distancia en la dirección x_2
- por lo tanto la variabilidad de x_1 es mayor que la de x_2

- son mas comunes coordenadas grandes en x_1 que en x_2
- tiene sentido ponderar mas a x_2 que a x_1

para construir la distancia estadística ...

1. $x_1^* = \frac{x}{\sqrt{s_{11}}}$ y $x_2^* = \frac{x_2}{\sqrt{S_{22}}}$ para ponderar las coordenadas con menor variabilidad
2. aplicamos la distancia euclideana para definir la distancia entre el punto $P = (x_1, x_2)$ al origen $O = (0, 0)$
3. si la variabilidad de ambas distancias es igual, se usa la distancia euclideana

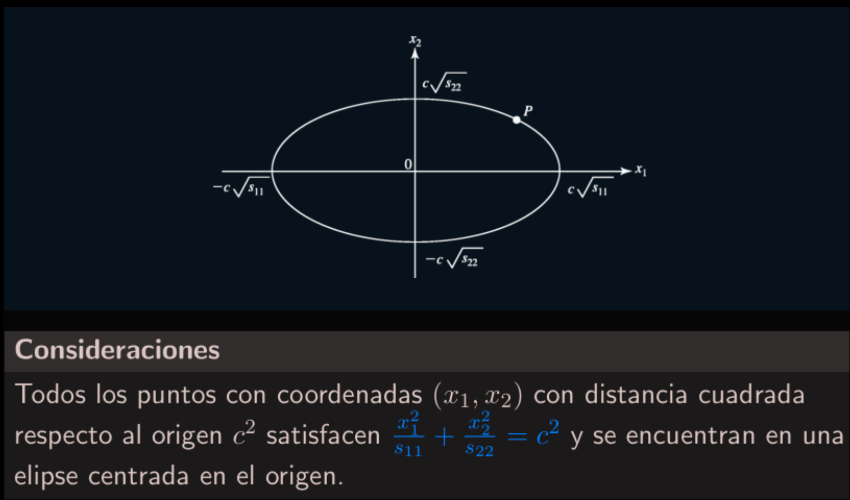


Figura 1: Consideraciones

por lo tanto la distancia estadística está dada como :

$$D_{estadistica}(P, Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + ... + \frac{x_{pp} - y_p{}^2}{s_{pp}}} \tag{2}$$

- todos los puntos P están a una distancia cuadrada constante del punto Q yacen sobre una hiper elipsoide centrada en Q con ejes mayor y menor paralelos a los ejes coordenados
- la distancia desde P hasta el origen O se puede obtener mediante $y_1 = y_2 = y_3 ... y_p = 0$
- si $s_{11} = s_{22}$ da igual que usar la distancia euclideana
- la distancia estadistica supone que las coordenadas son independientes
- no incluye los casos importantes

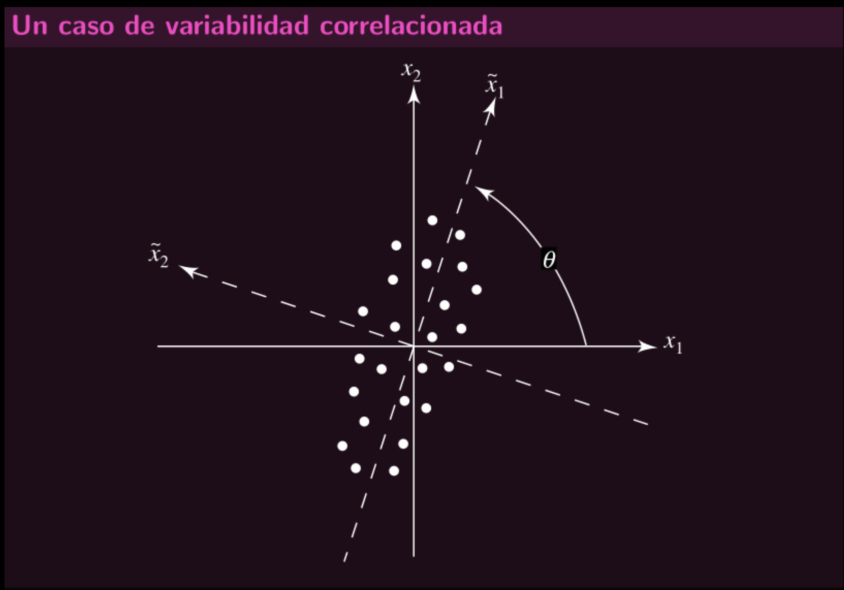


Figura 2: Caso

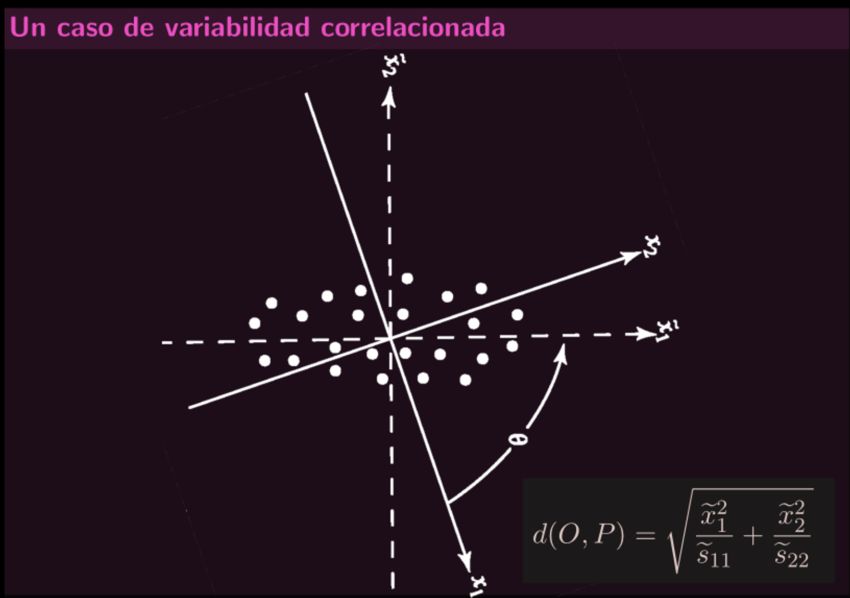


Figura 3: Caso2

cuando las variables están correlacionadas , la variable a un punto fijo es ...

$$D(P,Q)=\sqrt{a_{11}(x_1-y_1)^2+2a_{12}(x_1-y_1)(x_2-y_2)} \tag{3}$$

esta distancia puede ser calculada una vez se cauclulen a_{11}, a_{12}, a_{22}

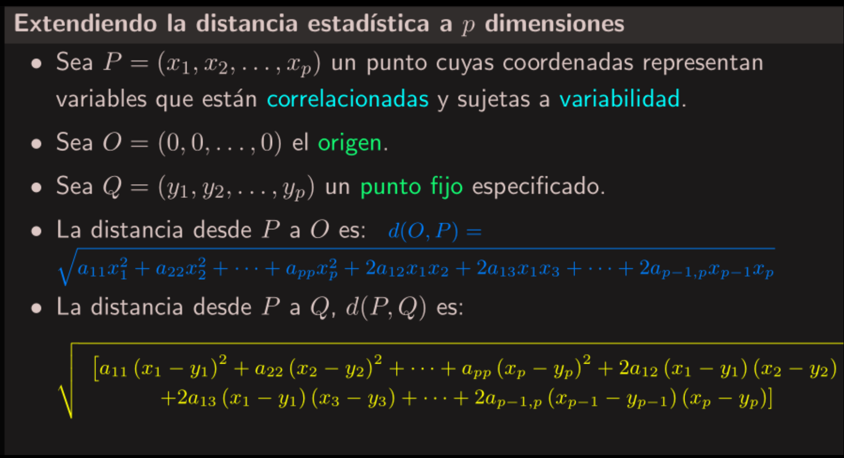


Figura 4: P-Dimensiones

Ahora, de forma matricial la distancia queda como :

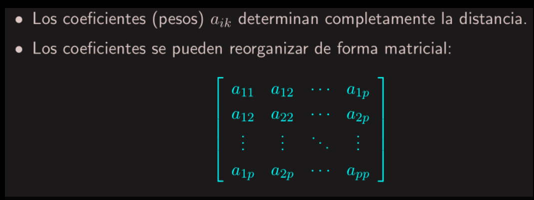


Figura 5: Matricial

3. cosas de repaso

producto interno

$$x'y = x_1y_1 + x_2y_2 + ... + x_ny_n \tag{4}$$

longitud de un vector

$$L_x = \sqrt{x_1^2 + x_2^2 + ... + x_n^2} = \sqrt{x'x} \tag{5}$$

angulo entre dos vectores ...

$$\cos\theta = \frac{x'y}{L_xL_y} = \frac{x'y}{\sqrt{x'x}\sqrt{y'y}} \tag{6}$$

independencia lineal
un conjunto de vectores $x_1, x_2, x_3, ..., x_n$ si existen constantes $c_1 ... c_n$ tales que $x_1c_1 + c_2x_2 + ... + c_kx_k = 0$

4. Dudas

1. no me queda claro la ecuacion 2 de donde sale la ecuacion 3