

## Distancia de Mahalanobis

Es una forma de medir la distancia entre sujetos , que a diferencia de la distancia euclidiana (Donde la frontera tiene una geometría circular ), esta tiene en cuenta las varianzas, de tal manera que clasifica mejor los outliers:

$$d_m(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T \Sigma^{-1} (\vec{x} - \vec{y})}.$$

Donde  $\Sigma^{-1}$  es la inversa de la matriz de covarianza entre los sujetos X y Y. Es una matriz cuadrada formada por covarianzas entre los dos sujetos

$$\Sigma = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Var}(X_n) \end{bmatrix}$$

la geometría de las curvas de nivel en la distancia de Mahalanobis es elíptica, de tal manera que la orientación de los ejes esta dado por la correlación entre las variables.

Se utiliza para detectar outliers y para cloustering.

## AIC

Los criterios de información abordan la relación entre la **complejidad** del modelo y la capacidad **predictiva** de un modelo. Estos mismos tienen la forma:

$$xIC = \text{Complejidad} - \text{Bondad de ajuste}$$

Para AIC:

$$AIC = 2k - 2 \ln(L)$$

Donde, k: numero de parámetros del modelo.

L: Máximo valor de la función de máxima verosimilitud.