

## Entrega de resúmenes

*Sebastián Marín Ruiz*

### Primera tarea: distancia de Mahalanobis

La distancia de Mahalanobis mide la distancia entre dos puntos en un espacio multivariable y se aplica cuando la distancia euclidiana se vuelve un método erróneo de medición. Para variables no correlacionadas, la distancia de Mahalanobis es igual a la distancia euclidiana. Sin embargo, cuando hay correlación entre dos o más variables, es necesario utilizar la distancia de Mahalanobis.

La distancia de Mahalanobis mide la distancia relativa al centroide, el cual es una base o punto central que puede ser considerado como un promedio general para los datos multivariable. Esta distancia se usa, principalmente, para encontrar outliers.

Formalmente, la distancia de Mahalanobis está definida como:

$$DM = [(X_B - X_A)^T * C^{-1} * (X_B - X_A)]^{0.5}$$

Donde  $X_A$  y  $X_B$  son un par de objetos y  $C$  es la matriz de covarianza muestral.

Una de las desventajas de esta técnica de medición es que la inversa de la matriz de correlación es necesaria para el cálculo, la cual no puede ser calculada para variables con una correlación muy alta.

### Segunda tarea: El Criterio de Información Bayesiano

El criterio de información bayesiano es un índice utilizado en la estadística bayesiana para elegir entre dos o más modelos alternativos.

Formalmente, está definido de la siguiente manera:

$$k \log(n) - 2 \log(L(\theta))$$

Donde  $n$  es el tamaño de la muestra, el número de observaciones o el número de puntos de datos con los que se esté trabajando;  $k$  es el número de parámetros estimados por el modelo y  $\theta$  es el conjunto de todos los parámetros.  $L(\theta)$  es la probabilidad del modelo utilizado, es decir, la probabilidad de obtener los datos que se tienen, suponiendo que el modelo que se está poniendo a prueba fuera dado.