



# Resumen: Distancia Mahalanobis

Universidad del Quindío  
Facultad de Ciencias Básicas y Tecnologías  
Programa de Física

Daniel Esteban Salazar Camacho

## Distancia Mahalanobis

La distancia Mahalanobis es una medida de la distancia entre un punto  $\mathbf{X}$  y una distribución de puntos  $Q$ . La idea consiste en medir a cuántas distribuciones estándar está  $\mathbf{X}$  de la media de  $Q$ . Esta distancia tiene en cuenta las correlaciones de las variables, no tiene unidades y no varía en escala.

Para calcular la distancia de Mahalanobis, primero se calcula la matriz de covarianza de los datos. Luego se computa la diferencia entre el punto y la media de la distribución, y se multiplica dicha diferencia por la inversa de la matriz de covarianza. Finalmente, se toma la raíz cuadrada de este producto. El valor resultante será la distancia entre el punto y la distribución. Los puntos que están más cerca de la distribución tendrán una distancia de Mahalanobis más pequeña, mientras que los puntos que están más lejos tendrán una distancia mayor.

De forma matemática, dada una distribución de puntos en  $\mathbb{R}^N$  con una media  $\bar{\mathbf{X}}$  y una matriz de covarianza  $\mathbf{C}_x$ , la distancia de Mahalanobis para un punto  $\mathbf{X}$  en dicha distribución, que es la matriz de datos que contiene  $n$  objetos en las filas medidas para  $p$  variables, es:

$$MD_i = \sqrt{(\mathbf{x}_i - \bar{\mathbf{x}})\mathbf{C}^{-1}_x(\mathbf{x}_i - \bar{\mathbf{x}})^T}$$

donde  $\mathbf{C}_x$  está dado por:

$$\mathbf{C}_x = \frac{1}{(n-1)}(\mathbf{X}_c)^T(\mathbf{X}_c) ;$$

siendo  $\mathbf{X}_c$  la matriz de datos centrada en columnas  $(\mathbf{X} - \bar{\mathbf{X}})$ .

La distancia Mahalanobis es usada regularmente en análisis multivariados: análisis de conglomerados, técnicas de clasificación, análisis de discriminantes, entre otros, debido a que es una medida de la similitud entre dos conjuntos de datos. También, dado que los outliers tienen las distancias más grandes, la distancia de Mahalanobis es útil para identificar dichos

puntos.

Otra de sus aplicaciones se encuentra en la clasificación de imágenes o señales de audio cuando se usa la distancia Mahalanobis en algoritmos de clasificación para determinar la similitud entre diferentes patrones.