

Cáncer de Mama

Cintha Merary Arias Mojica AM19011

2024-05-06

Método de Clasificación (Discriminate)

- ¿Las variables estaban altamente correlacionadas? ¿Por qué?
- ¿Para aplicar los métodos de clasificación necesitamos estandarizar los datos? ¿Por qué?

La respuesta es si aunque no en todos los casos (los algoritmos basados en arboles no suelen necesitarlo); pero especialmente en este caso si debemos debido a que vamos a usar un algoritmo basado en distancias y justamente las diferencias en escala pueden confundir al algoritmo y darle mas peso a algunas variables que otras por ese simple hecho; y eso es algo que deseamos evitar en nuestros análisis.

Justificación

1. Detección temprana y pronóstico: El análisis discriminante puede ayudar en la detección temprana del cáncer, lo que aumenta las posibilidades de tratamiento exitoso y supervivencia. Además, al clasificar con precisión los tipos de cáncer, puede proporcionar información sobre el pronóstico del paciente, lo que ayuda a los médicos a desarrollar planes de tratamiento personalizados.
2. Reducción de costos de atención médica: Al permitir un diagnóstico más preciso y oportuno, el análisis discriminante puede ayudar a reducir los costos asociados con pruebas médicas adicionales, tratamientos incorrectos y hospitalizaciones prolongadas.
3. Mejora de la calidad de vida: Una clasificación precisa del tipo de cáncer puede llevar a tratamientos más efectivos y menos invasivos, lo que puede mejorar la calidad de vida de los pacientes al reducir los efectos secundarios y las complicaciones asociadas con tratamientos agresivos innecesarios.
4. Equidad en la atención médica: Al proporcionar un método objetivo y estandarizado para clasificar los tipos de cáncer, el análisis discriminante puede ayudar a garantizar una atención médica más equitativa y consistente para todos los pacientes, independientemente de su ubicación geográfica o recursos disponibles.
5. Investigación y desarrollo de nuevas terapias: El análisis discriminante también puede tener un impacto en la investigación del cáncer al identificar subgrupos de pacientes con características biológicas similares que podrían beneficiarse de terapias específicas o participar en ensayos clínicos para el desarrollo de nuevos tratamientos.

Base de datos:

Los datos están formados por las siguientes variables:

Información categórica:

- 1) ID number
- 2) Diagnosis (M = maligno B = benigno)

Y luego se miden 10 características de las células que se observan en la imagen:

- a) radius
- b) texture
- c) perimeter
- d) area
- e) smoothness
- f) compactness
- g) concavity
- h) concave points
- i) symmetry
- j) fractal dimension

Y para cada una de estas características se miden 3 estimaciones: la media, la desviación típica y el máximo. Con lo cual tendremos 30 variables numéricas.

La variable Diagnosis contiene la clasificación de los cánceres en 357 benignos y 212 malignos.