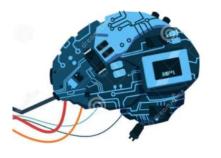
Big Data, Machine Learning & Business Intelligence

Por: Carlos Carreño

ccarreno@cienciadedatos.es



Regresión Logística con R

Introducción

Regresión Logística: Variables



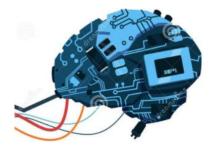
- Variable dependiente
 - Variable Dicotómica (toma solo dos valores). Ejemplo: TRUE, FALSE, Enfermo, Sano, Acepta o Rechaza.
- Variables Independientes
 - Numéricas o categóricas

Regresión Logística: Aplicaciones



- Estimación de probabilidad de un evento, en función de variables independientes
 - Probabilidad de estar empleado. (TRUE, FALSE) en función de profesión, nivel de ingles, estado civil, expectativa salarial.
 - Probabilidad de comprar un producto, en función de edad, genero, nivel socioeconómico, etc.
 - Probabilidad de pagar el crédito (TRUE, FALSE) en función de edad, estado civil, ingresos

Modelo de Regresión Logística



 La probabilidad de que la variable dependiente pertenezca a una categoría será una combinación lineal de las variables independientes

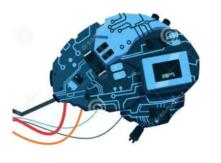
$$z_i = \ln(rac{p_i}{1-p_i}) = eta_0 + eta_1 x_1 + \ldots + eta_p x_p$$

Sera igual a:

$$p_i = 1 - \frac{1}{1 + \exp(z_i)}$$

• Si la probabilidad p_i es mayor a 0.5 se le asignara una categoría y si es menor la otra.

Framingham Risk Score



- La puntuación de riesgo de Framingham es un algoritmo específico de género que se utiliza para estimar el riesgo cardiovascular de un individuo a 10 años.
- Los sistemas de puntuación de riesgo cardiovascular dan una estimación de la probabilidad de que una persona desarrolle una enfermedad cardiovascular dentro de un período de tiempo específico, generalmente de 10 a 30 años

Framingham Risk Score: Variables



male

0 = Female; 1 = Male

age

Age at exam time.

education

1 = Some High School; 2 = High School or GED; 3 = Some College or Vocational School; 4 = college

currentSmoker

0 = nonsmoker; 1 = smoker

cigsPerDay

number of cigarettes smoked per day (estimated average)





BPMeds

0 = Not on Blood Pressure medications; 1 = Is on Blood Pressure medications

prevalentStroke

prevalentHyp

diabetes

0 = No; 1 = Yes

totChol

mg/dL

sysBP

mmHg

diaBP

mmHg





BMI

Body Mass Index calculated as: Weight (kg) / Height(meter-squared)

heartRate

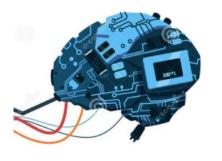
Beats/Min (Ventricular)

glucose

mg/dL

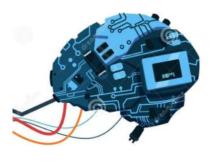
TenYearCHD

Laboratorio



 Usando el conjunto de datos del sistema de puntuación de Framigham, crear un modelo para predecir si una persona tendrá una enfermedad cardiovascular en un periodo de 10 años. Variable dependiente **TenYearCHD**

Referencias



• https://dtellogaete.medium.com/regresi%C3%B3n-log%C3%ADstica-en-python-y-r-machine-learning-02-fa066b3add09