

NOTACIÓN MODELIZACIÓN

Training data

FEATURES

x_1 x_2 x_3

ALTURA PESO N° PIE

PERSONA 1 → $x_1^{(1)}$ $x_2^{(1)}$ $x_3^{(1)}$

PERSONA 2 → $x_1^{(2)}$ $x_2^{(2)}$ $x_3^{(2)}$

PERSONA 3 → $x_1^{(3)}$ $x_2^{(3)}$ $x_3^{(3)}$

$$X = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} \\ x_1^{(3)} & x_2^{(3)} & x_3^{(3)} \end{pmatrix}$$

$$\vec{x}^{(i)} = (x_1^{(i)} \ x_2^{(i)} \ x_3^{(i)})$$

Es un dataframe

TARGET

y

SEXO

$y^{(1)}$

$y^{(2)}$

$y^{(3)}$

$$y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \end{pmatrix}$$

MODELO

Test data

Es otra fila que añadimos al DataFrame

x_1

x_2

x_3

y

ALTURA

PESO

N° PIE

SEXO

PERSONA NUEVA:

x_1

x_2

x_3

¿?

Estas la y que tengo que predecir

NOTACIÓN / RESUMEN

MODELIZACIÓN

EN GENERAL:

Training data: Es un DataFrame que denoto como una matriz:

$$\left(\begin{array}{c|c} X & y \end{array} \right)$$

↑ FEATURES ↑ TARGET

Matriz $d \times n$

$$X = \begin{pmatrix} \vec{x}^{(1)} \\ \vdots \\ \vec{x}^{(n)} \end{pmatrix} = \begin{pmatrix} x_1^{(1)} & \dots & x_d^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(n)} & \dots & x_d^{(n)} \end{pmatrix}$$

$\vec{x}^{(i)} = (x_1^{(i)} \dots x_d^{(i)})$
← Vector fila $d \times 1$

$y = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(n)} \end{pmatrix}$ ← Vector columna $1 \times n$

$n \equiv$ Número de filas (entregas de la base de datos)

$d \equiv$ Número de features

Utilizo el training data para construir un modelo

[DEFINIR + AJUSTAR]

Puedo pensar en un modelo como una función f que para cada x predice un valor de y .

Test data: Utilizo datos nuevos para probar el modelo. Son como nuevas filas del dataframe, pero sin la variable target.

m DATOS NUEVOS

$$\rightarrow X = \begin{pmatrix} \vec{x}^{(1)} \\ \vdots \\ \vec{x}^{(m)} \end{pmatrix} \quad ? y?$$
$$\hat{y} = f(X) = \begin{pmatrix} f(\vec{x}^{(1)}) \\ \vdots \\ f(\vec{x}^{(m)}) \end{pmatrix}$$

[PREDECIR]

Utilizando el modelo predigo valores para y :

Si conozco los valores reales de y para estos nuevos datos, puedo compararlos con los predichos:

$$y \sim \hat{y} \rightarrow \text{PRECISIÓN DEL MODELO}$$

[EVALUAR]