

# Intelixencia Artificial

## [G4012328]

### P3.1. Práctica de Regresión Lineal

Curso 2023 / 2024

**3º Grao en Enx. Informática**

José María Alonso Moral

[Josemaria.alonso.moral@usc.es](mailto:Josemaria.alonso.moral@usc.es)

<https://citius.gal/team/jose-maria-alonso-moral>

Escola Técnica Superior de Enxeñaría (ETSE)

Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS)

Universidade de Santiago de Compostela (USC)

# Aprendizaje Supervisado: Problema de Regresión

## ❑ Aprender la función que mejor represente los datos disponibles

- **Datos tabulares etiquetados:** Entrada versus Salida
- **Validación:** Entrenamiento vs Test

## ❑ Notación

- $m$  = número de ejemplos de entrenamiento
- $x$  = valores de entrada / **características**
- $y$  = valor de salida / **respuesta**
  - $(x, y)$  –par entrada-salida genérico
  - $(x_i, y_i)$  –par entrada-salida  $i$ ésimo

## ❑ **MIMO:** Multiple-inputs, Multiple-outputs

## ❑ **MISO:** Multiple-inputs, Single-output

# Aprendizaje Supervisado: Problema de Regresión

## ❑ Función Lineal

$$y = h_{\theta}(x) = (\theta_0 + \theta_1 x) = Ax + B$$

Valor en el Origen

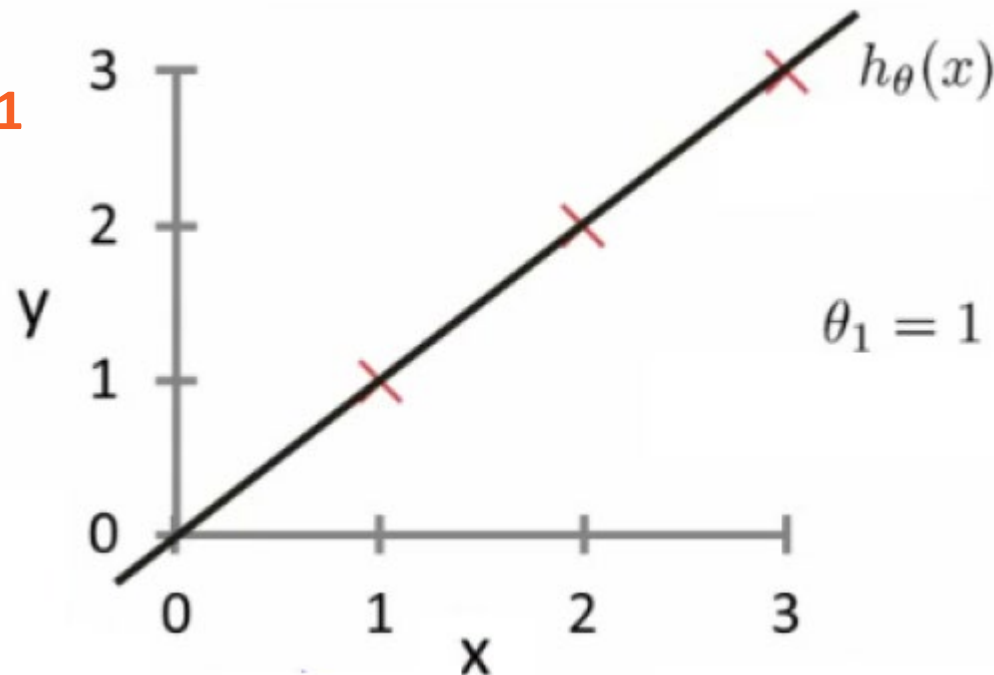
Gradiente

$$B = y - Ax$$

$$A = (y - B) / x$$

Si  $x=0$ ,  $y=B=0$

Si  $x=1$ ,  $y=A+B=1$ ,  $A=1$



# Aprendizaje Supervisado: Problema de Regresión

## □ Aprendizaje de la Función Lineal

### ➤ Guiado por una función de coste:

#### ❖ Absolute Error

$$\left| h_{\Theta}(x^{(i)}) - y^{(i)} \right|$$

#### ❖ MEA: Mean Absolute Error

#### ❖ Square Error

$$\left( h_{\Theta}(x^{(i)}) - y^{(i)} \right)^2$$

#### ❖ RMSE: Root Mean Square Error

### ➤ Objetivo de optimización: Minimizar el error

# Aprendizaje Supervisado: Problema de Regresión

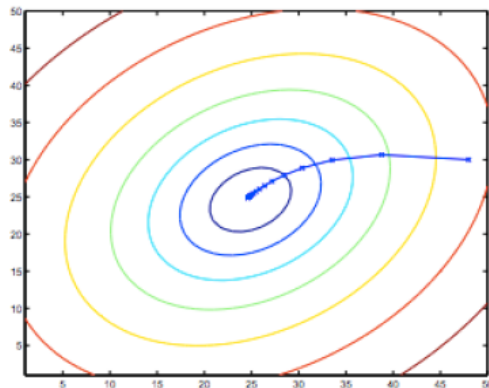
## □ Aprendizaje de la Función Lineal por Descenso de Gradiente

### ➤ Problema multivariable

$$h_0(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots + \theta_n x_n$$

$$h_0(x) = \theta^T X$$

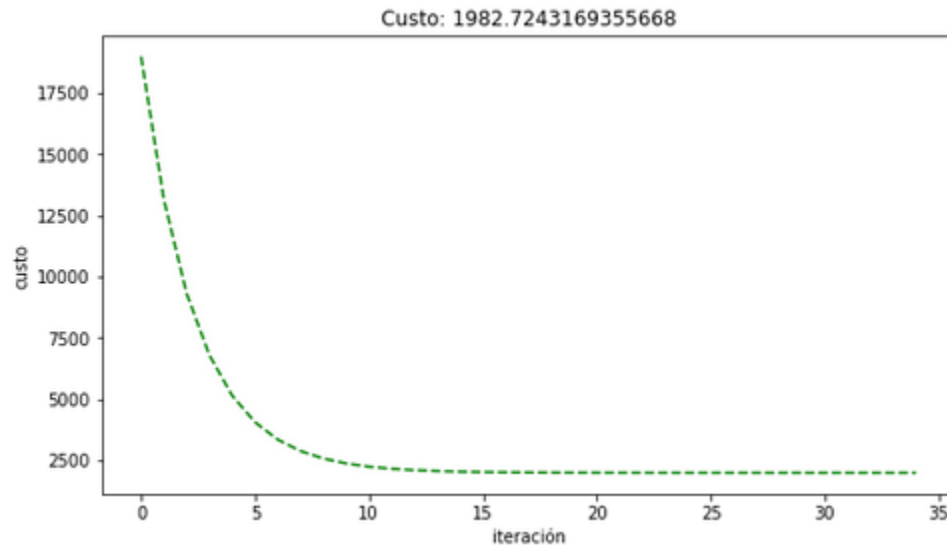
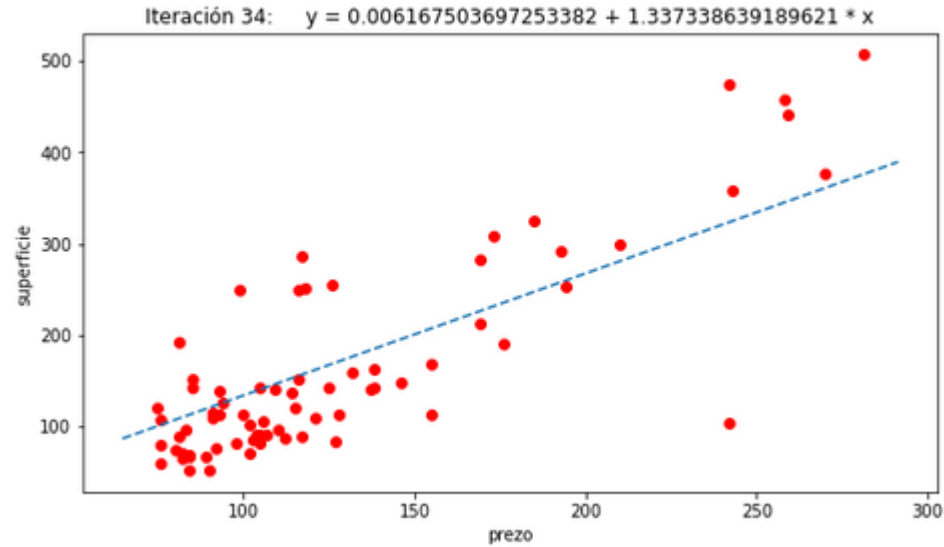
$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$



$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_{\theta}(x) - y)^2$$

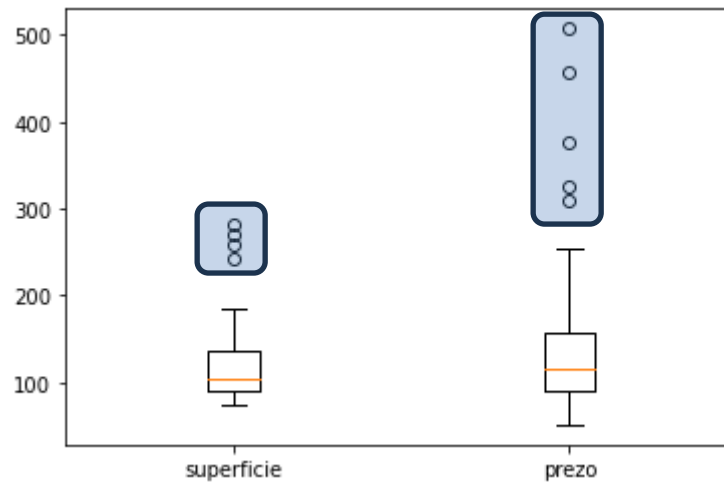
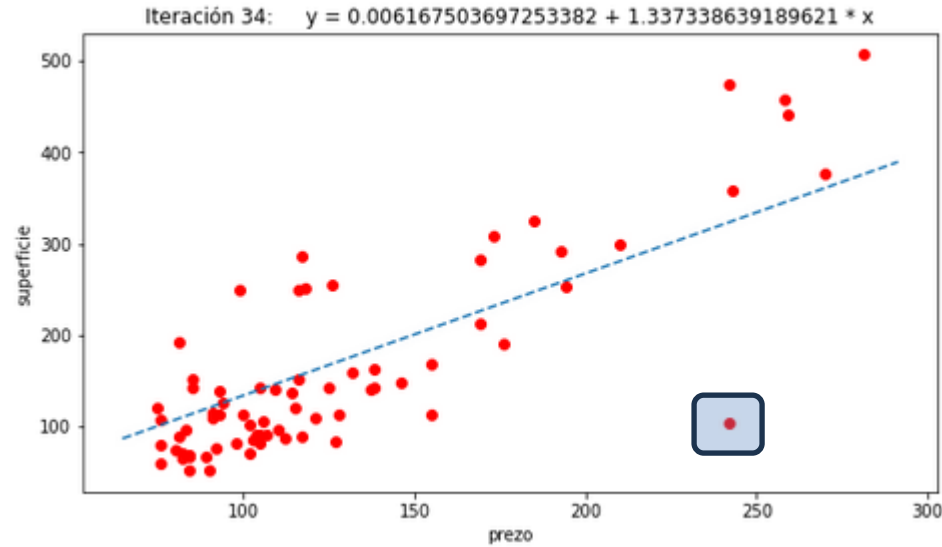
# Aprendizaje Supervisado: Ejemplo

## □ Viviendas



# Aprendizaje Supervisado: Ejemplo (Outliers)

## □ Viviendas



# Aprendizaje Supervisado: Problema de Cereales

<https://www.kaggle.com/datasets/crawford/80-cereals>

## Entradas

- **name:** Name of cereal
- **mfr** (Manufacturer of cereal)
  - A = American Home Food Products; G = General Mills; K = Kelloggs; N = Nabisco; P = Post; Q = Quaker Oats; R = Ralston Purina
- **type:** cold / hot
- **calories:** calories per serving
- **protein:** grams of protein
- **fat:** grams of fat
- **sodium:** milligrams of sodium
- **fiber:** grams of dietary fiber
- **carbo:** grams of complex carbohydrates
- **sugars:** grams of sugars
- **potass:** milligrams of potassium
- **vitamins:** vitamins and minerals - 0, 25, or 100, indicating the typical percentage of FDA recommended
- **shelf:** display shelf (1, 2, or 3, counting from the floor)
- **weight:** weight in ounces of one serving
- **cups:** number of cups in one serving

Construir un modelo de Regresión Lineal con 2 Entradas como variables predictoras, para predecir lo mejor posible la Salida “rating”

## Salida

- **rating:** a rating of the cereals (Possibly from Consumer Reports?)



# Referencias

- ❑ Kaggle (<https://www.kaggle.com/>)
- ❑ Stuart Russell and Peter Norvig, “**Artificial Intelligence: A Modern Approach**”, 4<sup>th</sup> US Ed.
  - <https://aima.cs.berkeley.edu/>
  - <https://github.com/aimacode/aima-python>
- ❑ **Aula Virtual:** “Aprendizaxe automática / Fundamentos e Regresión Lineal” (Senén Barro Ameneiro)