

Data Science & Machine Learning



red.es

Centro de
Bases de Datos
en Ciencia, Tecnología
y Planificación

GRN
Digital



Barrabés

The Valley

"El FSE invierte en tu futuro"
Fondo Social Europeo

Data Science & Machine Learning



Data Science & Machine Learning



GOBIERNO
DE ESPAÑA

VICEPRESIDENCIA
PRIMERA DEL GOBIERNO
MINISTERIO
DE ASUNTOS ECONÓMICOS
Y TRANSFORMACIÓN DIGITAL

SECRETARÍA DE ESTADO
DE DIGITALIZACIÓN
E INTELIGENCIA ARTIFICIAL

red.es

Centro de
Bases de Datos
en Ciencia, Tecnología
y Planificación

CRN
Digital

GARANTÍA
JUVENIL



UNIÓN EUROPEA

Barrabés

The Valley

"El FSE invierte en tu futuro"
Fondo Social Europeo

Índice

1. Regresión

1.1 Regresión Simple vs Regresión Multiple

1.2 Regresión Lineal vs Regresión No Lineal

1.2 Regresión Lineal Simple



2. Prophet



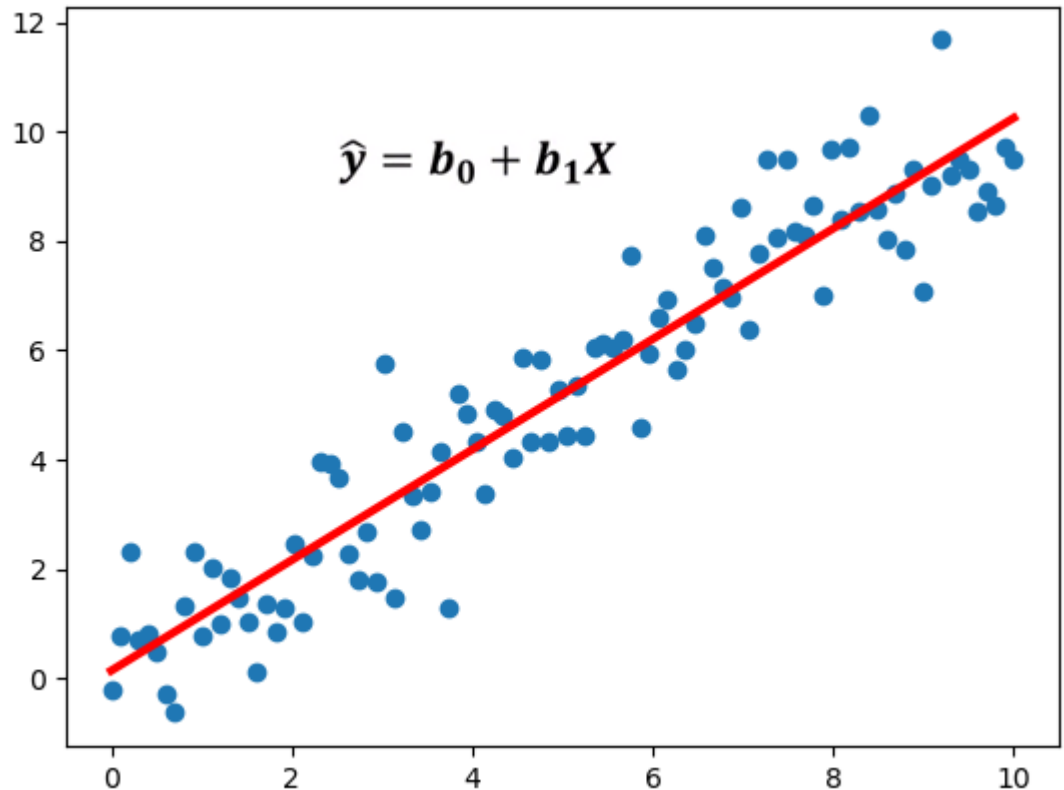
3. Neural Prophet

Regresión

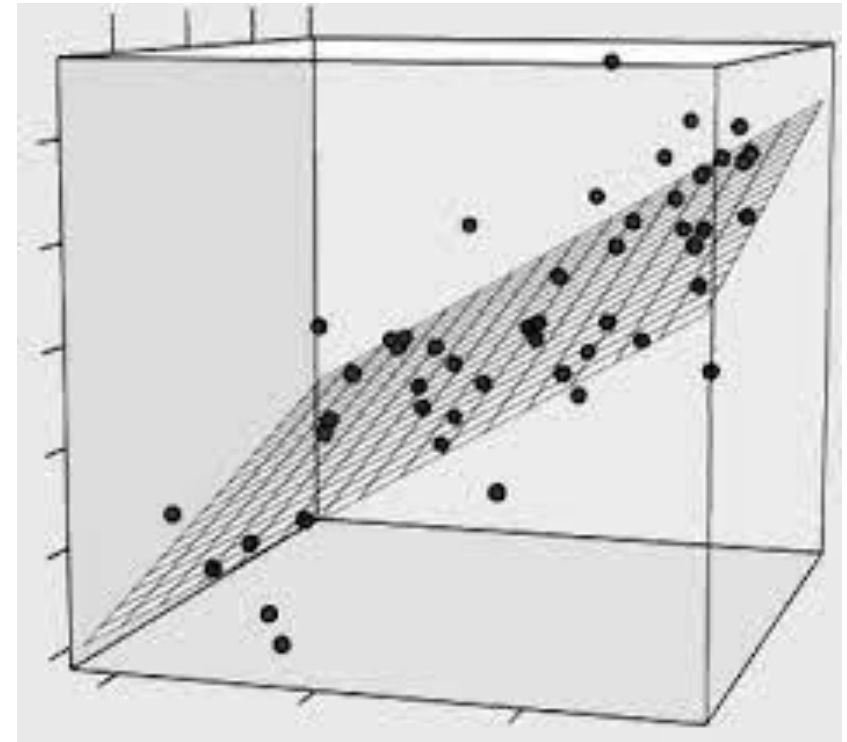
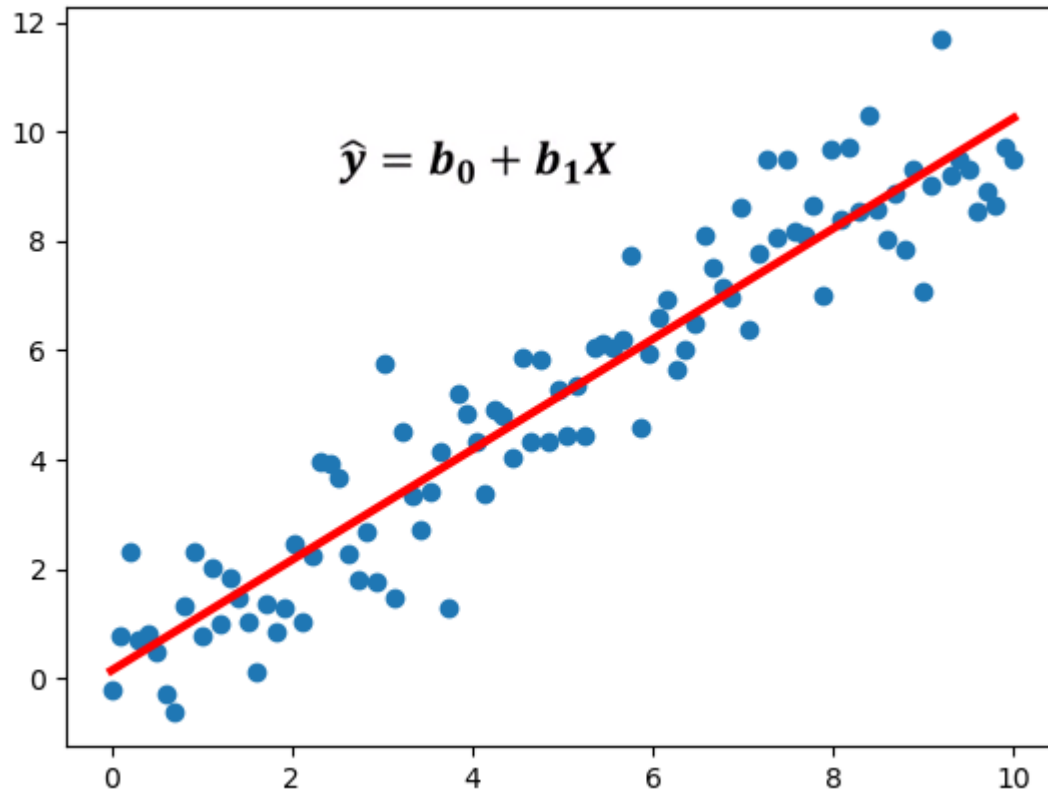
La regresión modela la relación entre una variable dependiente (Y) y una o más variables independientes (X)

Y es una variable continua

**Regresión vs Correlación:
La regresión permite hacer predicciones**

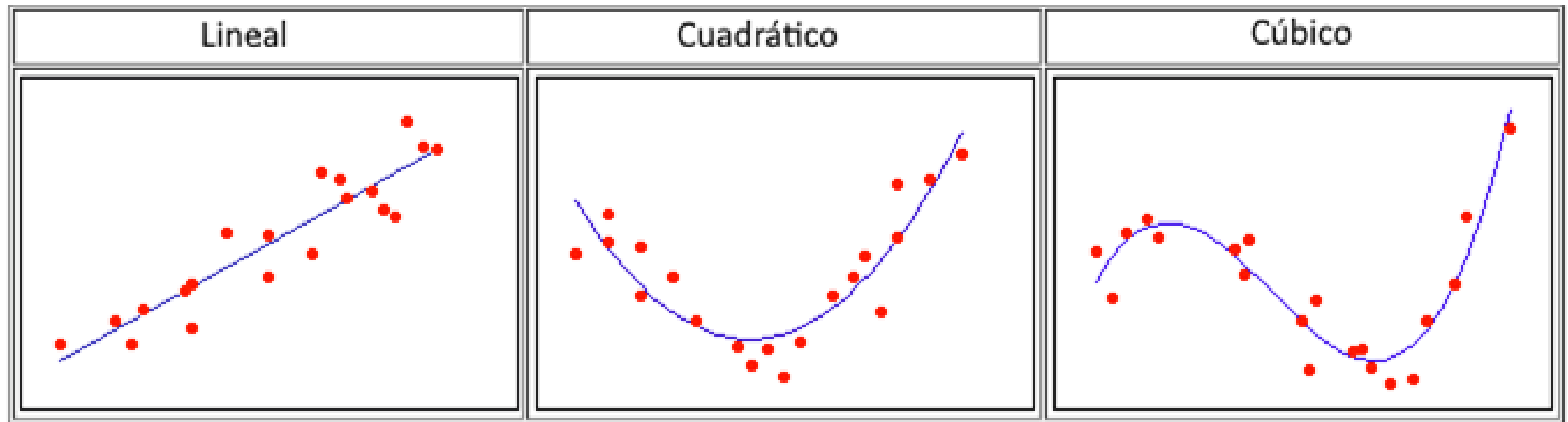


Regresión Simple vs Regresión Multiple

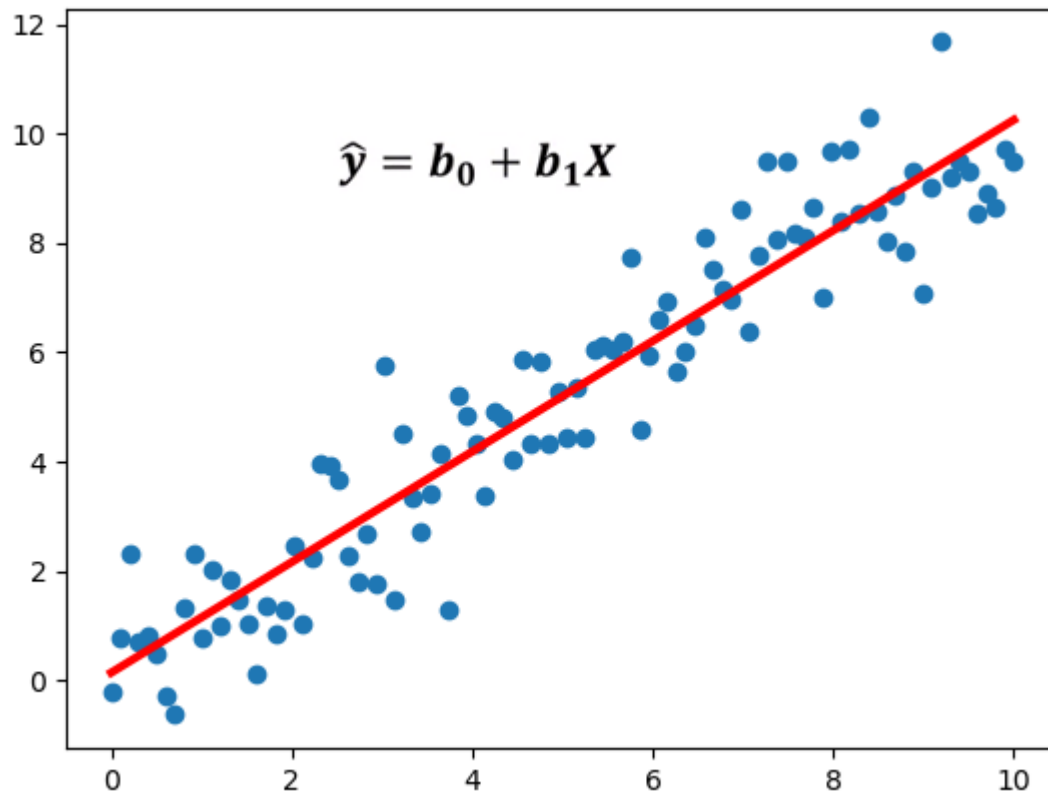


Valor de y	y-intercepto estimado	Pendientes estimadas		Error muestral
↓	↓	↓	↓	↓
$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki} + e_i$				

Regresión Lineal vs Regresión No Lineal

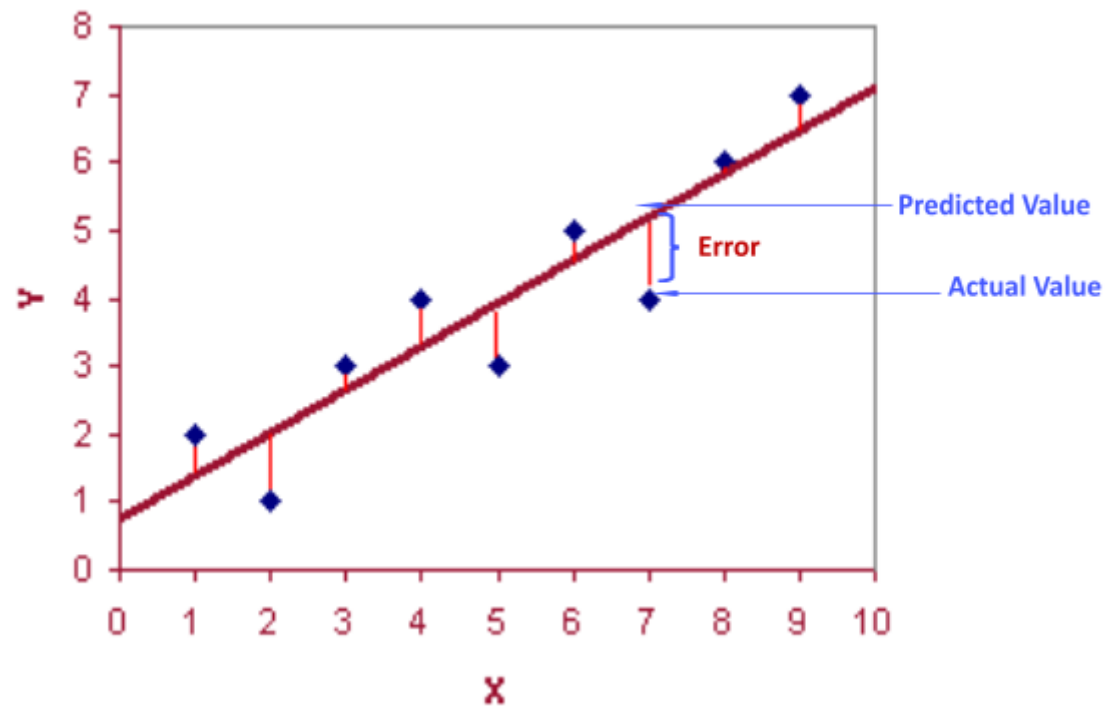


Regresión Lineal Simple



PREDICCIONES

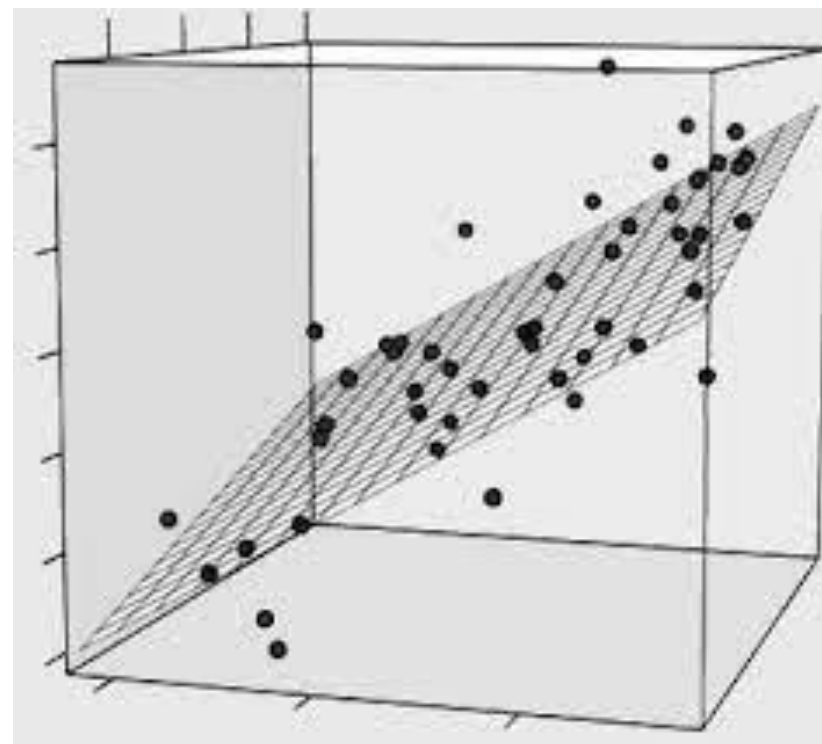
Regresión Lineal Simple



$$\text{Min}_{w,b} \text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \frac{1}{N} \sum_{i=1}^N (y_i - f_{w,b}(X_i))^2$$

Regresión Lineal Multiple

PREDICCIONES



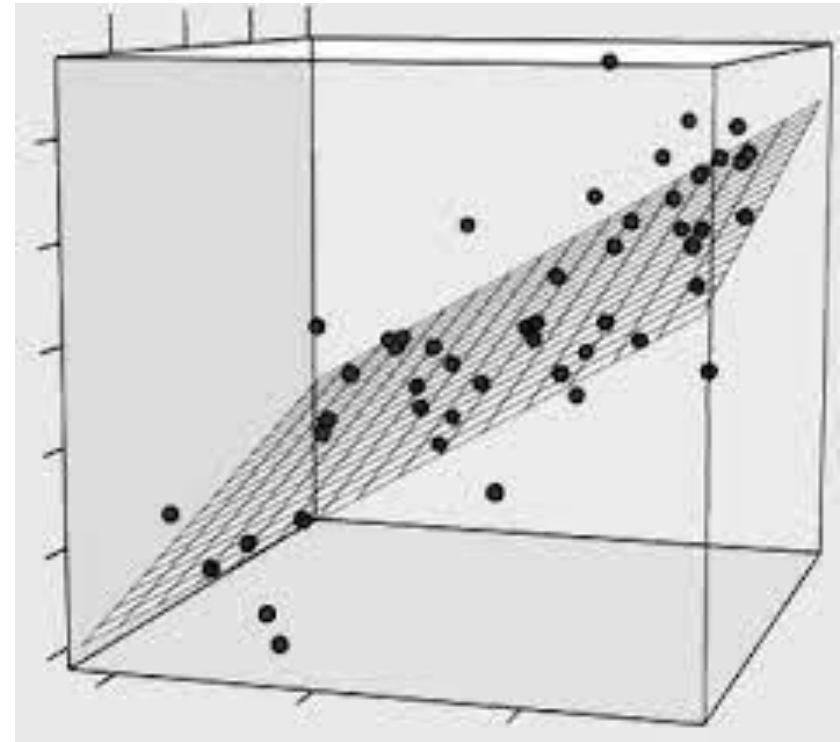
Valor de y y-intercepto estimado Pendientes estimadas Error muestral

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki} + e_i$$

15-7

Regresión Lineal Multiple

¡El problema de la colinealidad!



Valor de y	y -intercepto estimado	Pendientes estimadas		Error muestral
y_i	b_0	b_1	b_2	b_k
	$+ b_1 X_{1i}$	$+ b_2 X_{2i}$	$+ \dots + b_k X_{ki}$	$+ e_i$

$$y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki} + e_i$$

15-7

Regresión Lineal Multiple

¡El problema de la colinealidad!

Soluciones:
RIDGE
LASSO
ELASTIC-NET

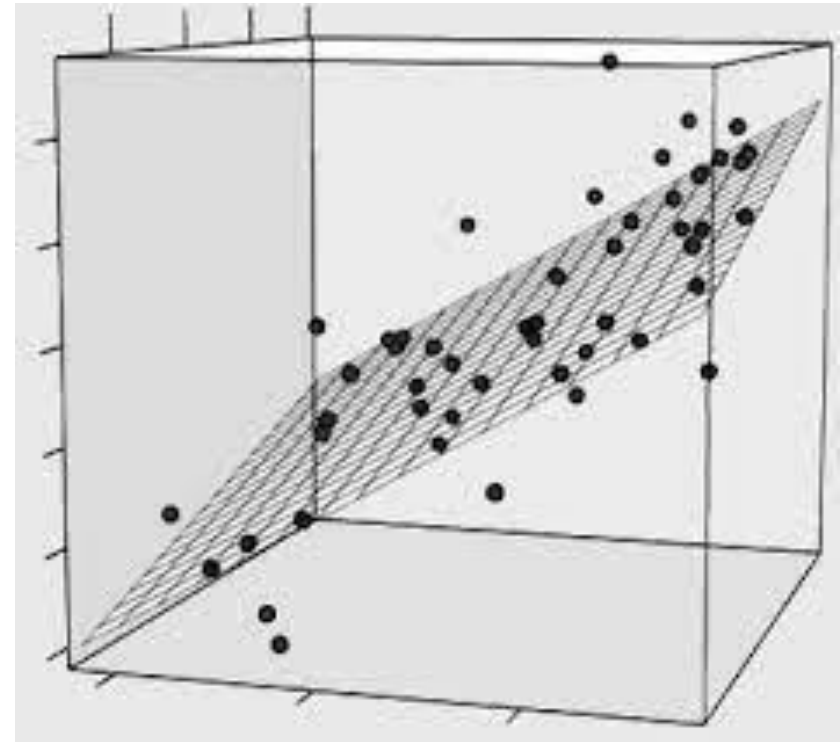


Diagram illustrating the components of the multiple linear regression equation:

$$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki} + e_i$$

Labels and their corresponding terms in the equation:

- Valor de y points to y_i
- y-intercept estimado points to b_0
- Pendientes estimadas points to b_1x_{1i} , b_2x_{2i} , and b_kx_{ki}
- Error muestral points to e_i

15-7

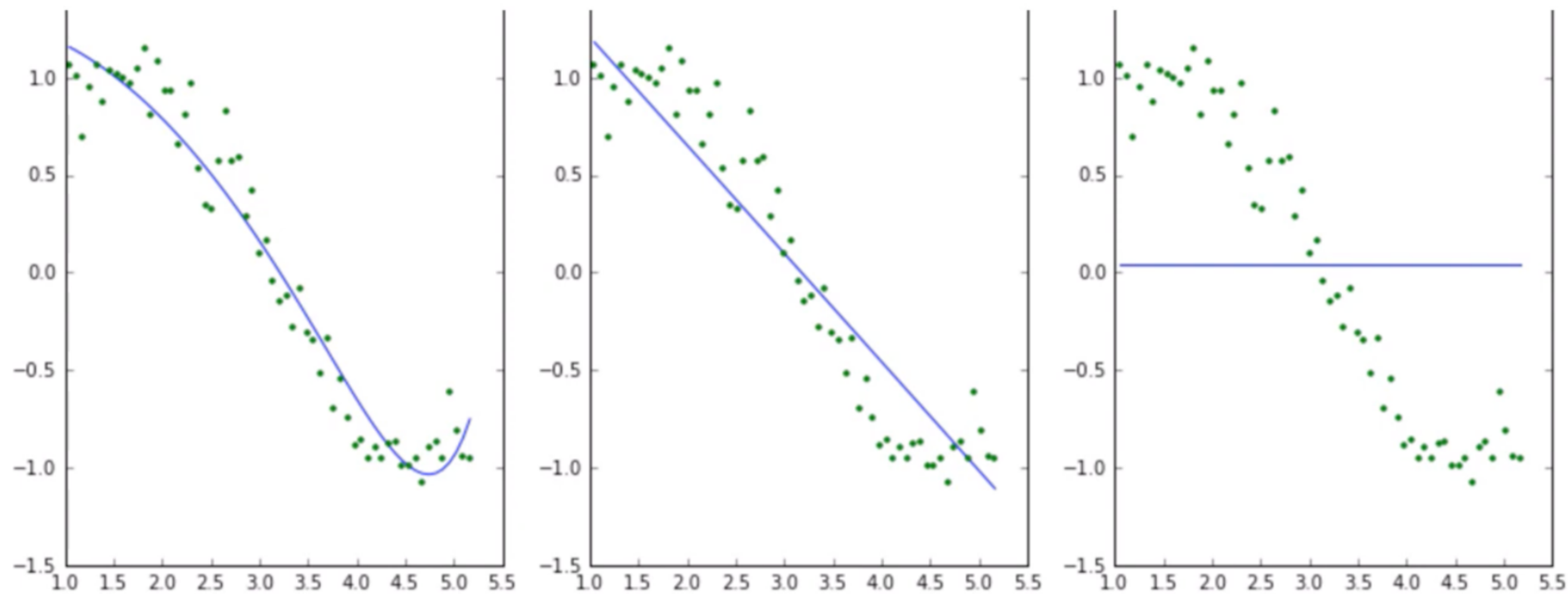
Regresión Lineal Multiple (RIDGE)

**Solucionamos la colinealidad
con una penalización al tamaño
de los coeficientes**

$$\text{Min} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f_{w,b}(X_i) \right)^2 + \lambda \sum_{j=1}^D w_j^2 \right]$$

Regresión Lineal Multiple (RIDGE)

Cuanto mayor es lambda más simple es el modelo



$$\text{Min} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f_{w,b}(X_i) \right)^2 + \lambda \sum_{j=1}^D w_j^2 \right]$$

Regresión Lineal Multiple (LASSO)

**Solucionamos la colinealidad
con una penalización al tamaño
de los coeficientes**

$$= \text{Min} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f_{\mathbf{w},b}(X_i) \right)^2 + \lambda \sum_{j=1}^D |w_j| \right]$$

Regresión Lineal Multiple (LASSO)

LASSO

$$= \text{Min} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f_{\mathbf{w},b}(X_i) \right)^2 + \lambda \sum_{j=1}^D |w_j| \right]$$

VS

RIDGE

$$= \text{Min} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f_{\mathbf{w},b}(X_i) \right)^2 + \lambda \sum_{j=1}^D w_j^2 \right]$$

Regresión Lineal Múltiple (ELASTICNET)

**Solucionamos la colinealidad
con una penalización al tamaño
de los coeficientes**

$$\text{Min} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f_{\mathbf{w},b}(X_i) \right)^2 + \lambda_1 \sum_{j=1}^D |w_j| + \lambda_2 \sum_{j=1}^D w_j^2 \right]$$

Regresión Lineal Multiple (ELASTIC NET)

LASSO

$$= \text{Min} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f_{w,b}(X_i) \right)^2 + \lambda \sum_{j=1}^D |w_j| \right]$$

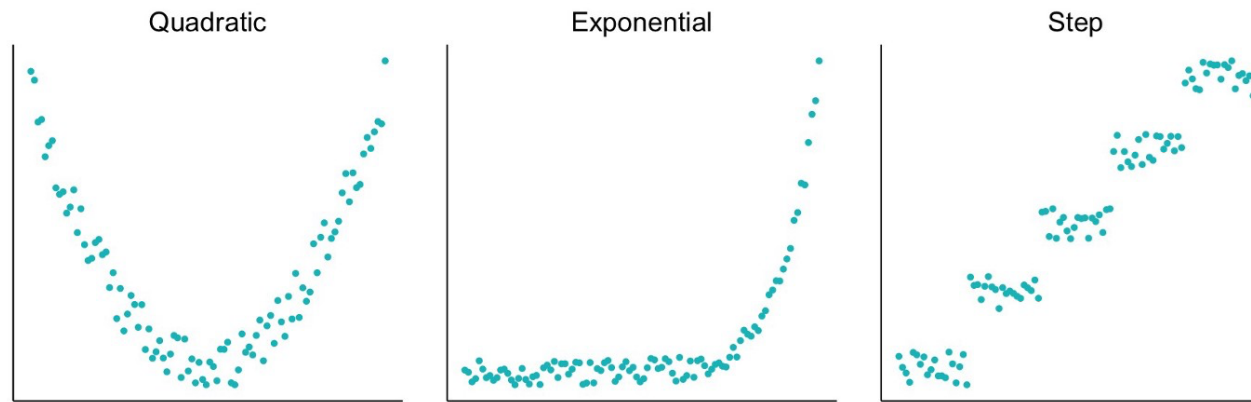
RIDGE

$$: \text{Min} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f_{w,b}(X_i) \right)^2 + \lambda \sum_{j=1}^D w_j^2 \right]$$

ELASTIC NET

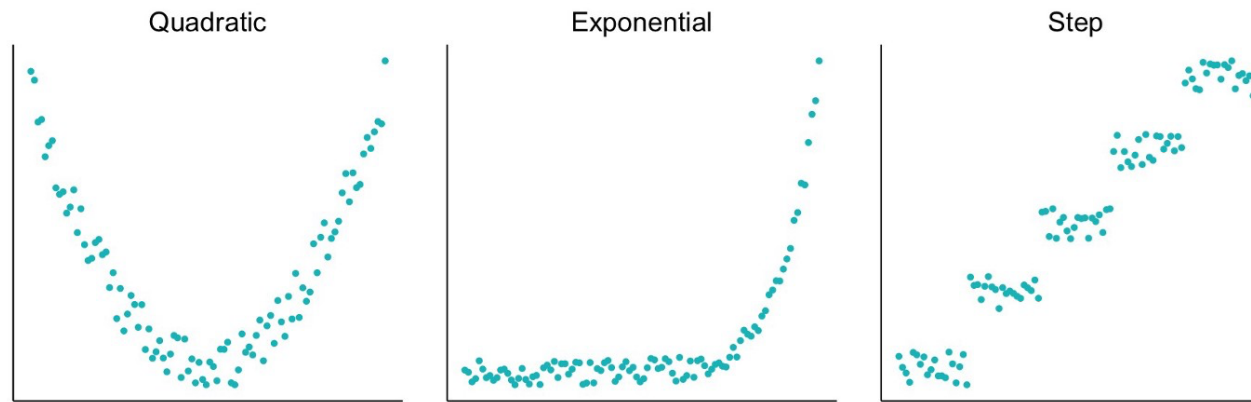
$$\text{Min} \left[\frac{1}{N} \sum_{i=1}^N \left(y_i - f_{w,b}(X_i) \right)^2 + \lambda_1 \sum_{j=1}^D |w_j| + \lambda_2 \sum_{j=1}^D w_j^2 \right]$$

Regresión no lineal



**¿Y SI NO SABEMOS LA FUNCIÓN QUE GENERA
LOS DATOS?**

Regresión no lineal

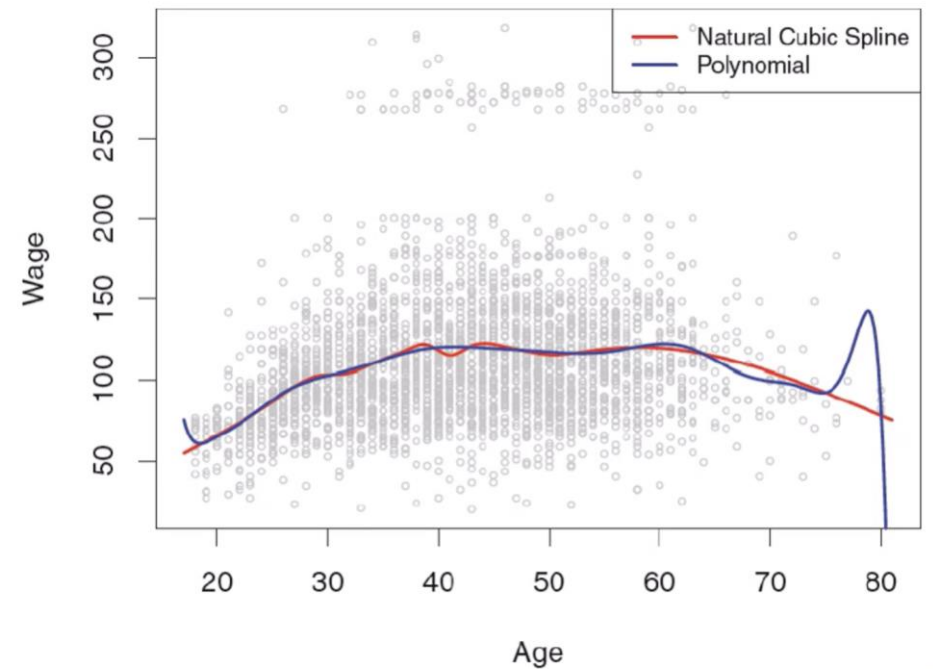


**¿Y SI NO SABEMOS LA FUNCIÓN QUE GENERA
LOS DATOS?**

**APROXIMAMOS CON POLINOMIOS O USAMOS
MÉTODOS GENÉRICOS**

Regresión no lineal (Polinómica)

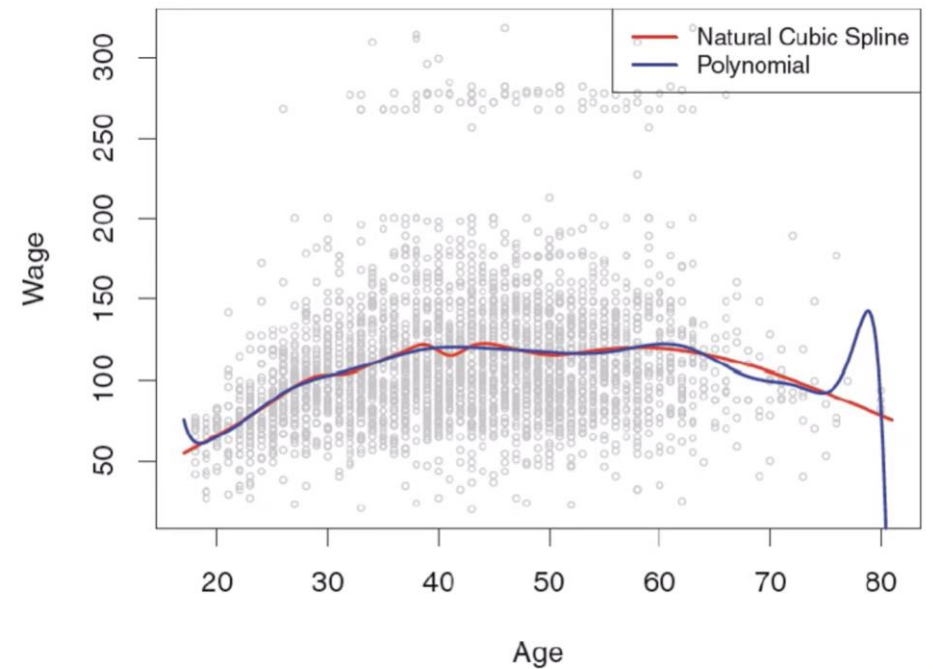
**COMPORTAMIENTO
EXTRAÑO EN LOS EXTREMOS**



Regresión no lineal (Polinómica)

**COMPORTAMIENTO
EXTRAÑO EN LOS EXTREMOS**

**MÉTODOS GENÉRICOS COMO:
KNN
DECISION TREES
RANDOM FOREST**

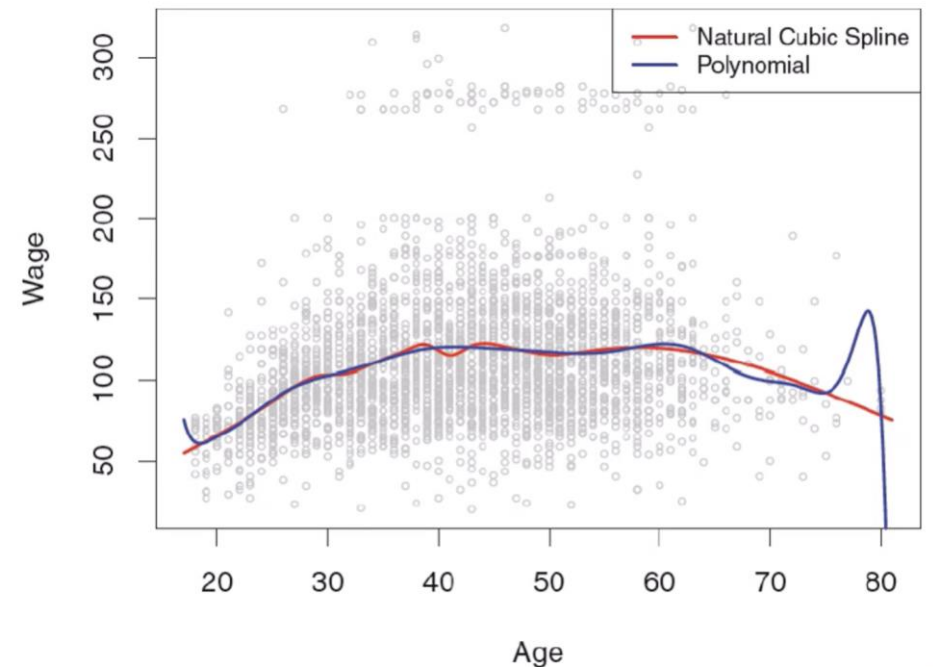


Regresión no lineal (Polinómica)

**COMPORTAMIENTO
EXTRAÑO EN LOS EXTREMOS**

**MÉTODOS GENÉRICOS COMO:
KNN
DECISION TREES
RANDOM FOREST**

**LOS ESTUDIAREMOS EN
CLASIFICACIÓN**



¿Qué es Prophet?



- Desarrollado por Facebook
- Open Source desde 2017
- Inicialmente desarrollado para predecir información sobre los usuarios de Facebook.
- Uno de sus objetivos principales era mejorar la predicción escalada.

Package Forecast



Vs.



Rob J Hyndman:

Developed the
forecast package

Author of the blog
'Hyndsight'

- ✓ En constante desarrollo
- ✓ Hace predicciones < 1 segundo
- ✓ Permite ajustar parámetros
- ✓ Rinde bien con datos incompletos
- ✓ Fácil de aplicar a un negocio

- ✓ En constante desarrollo
- X Hace predicciones ~2 minutes.
- ✓ Permite ajustar parámetros
- X Falla cuando tiene gaps en los datos
- X Difícil de escalar para un negocio

¿Y si necesitamos algo más potente?

Neural Prophet

```
1 from neuralprophet import NeuralProphet
2 import pandas as pd
3
4 df = pd.read_csv('toiletpaper_daily_sales.csv')
5
6 m = NeuralProphet()
7
8 metrics = m.fit(df, freq="D")
9
10 forecast = m.predict(df)
```




Contacto:

luis@lubay.es





red.es



UNIÓN EUROPEA

"El FSE invierte en tu futuro"

Fondo Social Europeo

