

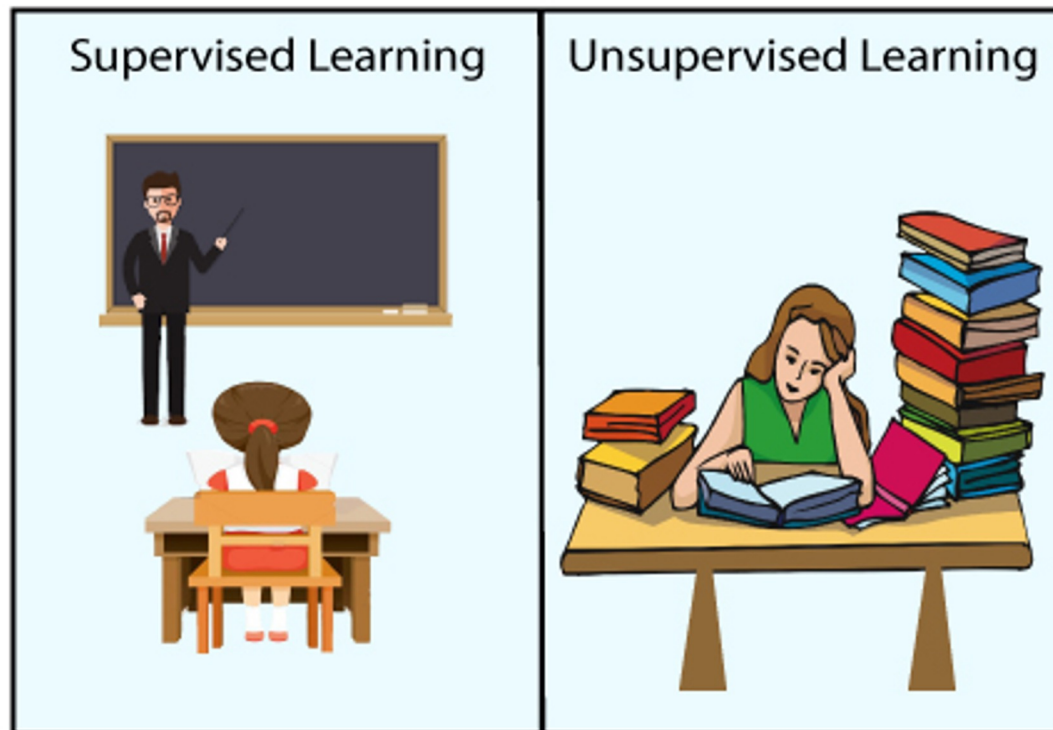
# Introducción al Aprendizaje Supervisado

- 1.- Qué es el aprendizaje supervisado.
- 2.- El problema de regresión.

Que es el aprendizaje  
supervisado.

# Aprendizaje supervisado

¿Qué es aprendizaje supervisado y ML?

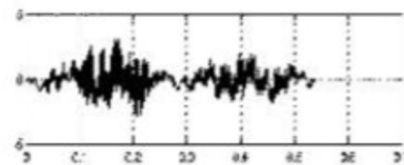
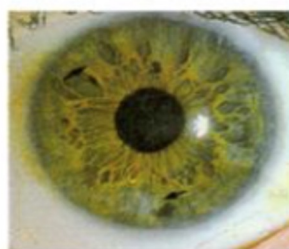


## ¿Qué es un patrón?

Es una abstracción, representada por un conjunto de mediciones que describen un objeto “físico”

Existen de diversos tipos:

Visuales, temporales, sónicos, lógicos, ...



John Smith

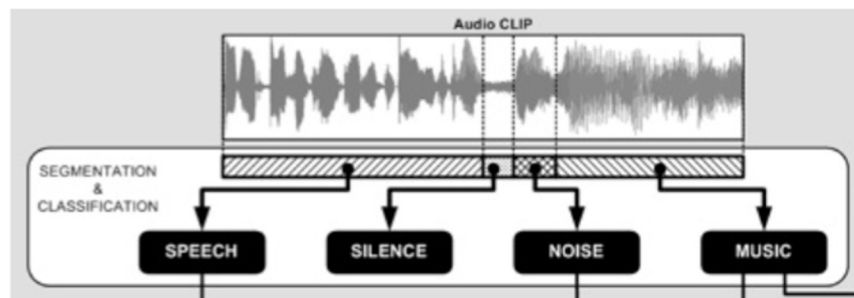
## ¿Qué es la *clase* de un patrón?

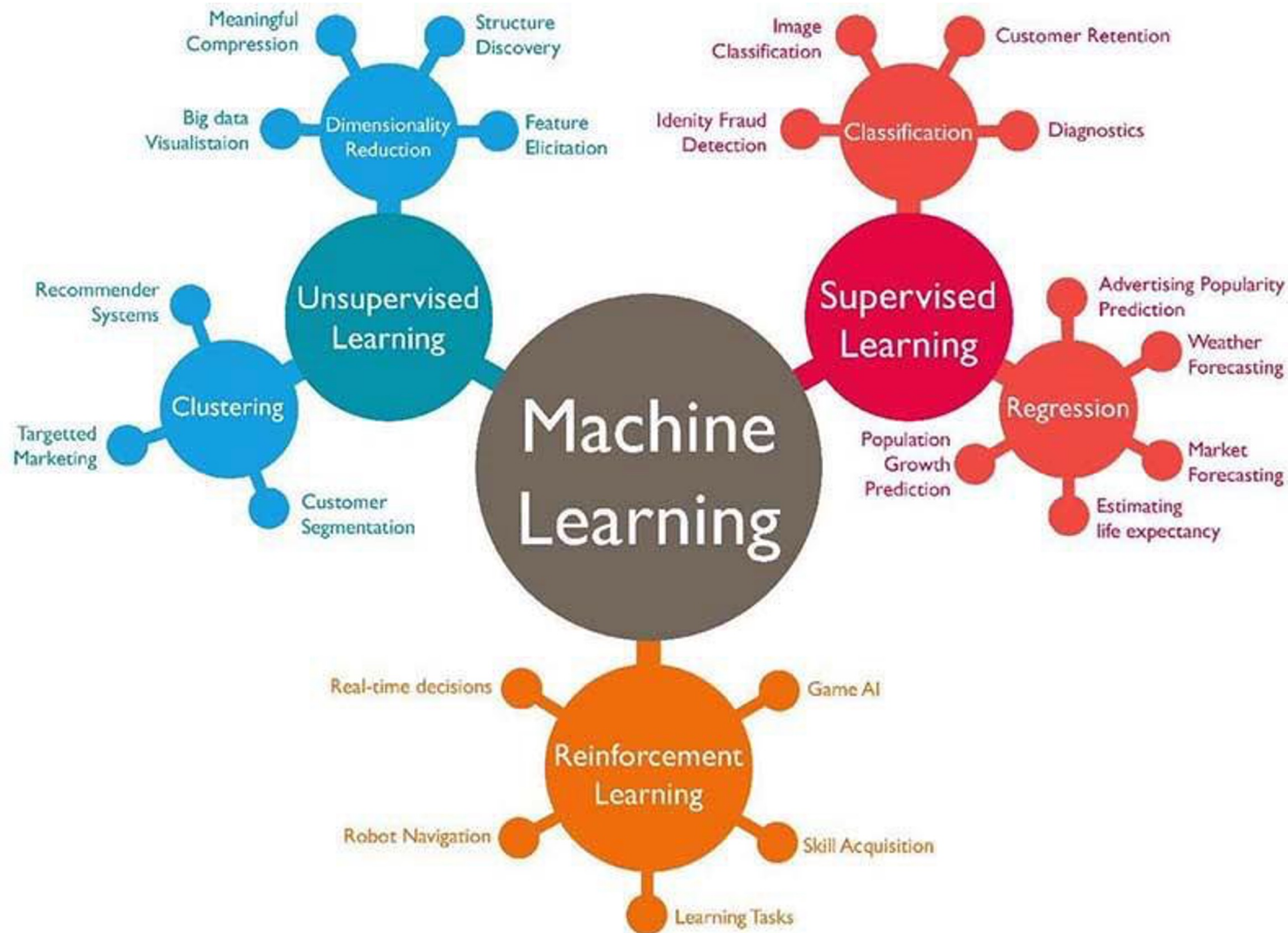
(también llamada *categoría* o *clase informacional*)

Es un conjunto de patrones que comparten características en común

Una colección de objetos “similares”, no necesariamente idénticos

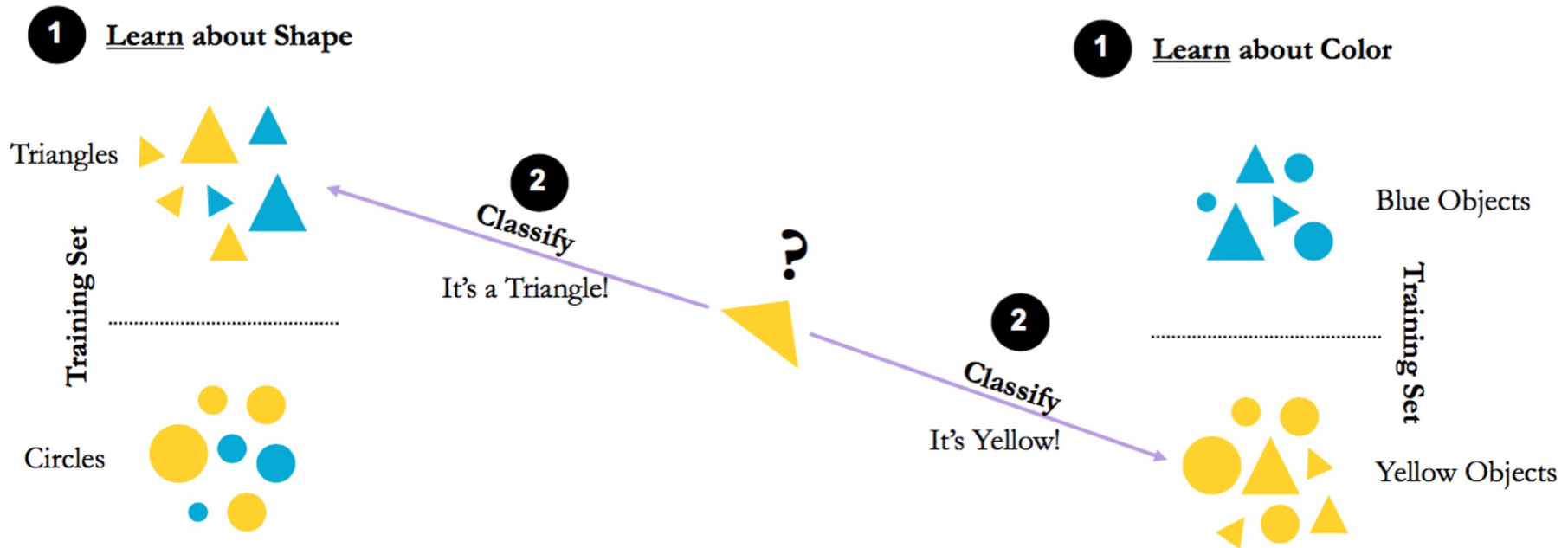
Al reconocer un patrón, se le asigna una clase prescrita.





## Aprendizaje Supervisado

Un maestro provee conjuntos de entrenamiento etiquetados, usados para entrenar un clasificador





Las columnas serán nuestras: variables (features en ingles) independientes.

Las filas serán:

- Ejemplos
- Sujetos
- Muestras
- etc.

User ID	Gender	Age	Salary	Purchased	Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
15624510	Male	19	19000	0	10.69261758	986.882019	54.19337313	195.7150879	3.278597116
15810944	Male	35	20000	1	13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
15668575	Female	26	43000	0	17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
15603246	Female	27	57000	0	20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
15804002	Male	19	76000	1	22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
15728773	Male	27	58000	1	24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
15598044	Female	27	84000	0	24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
15694829	Female	32	150000	1	23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
15600575	Male	25	33000	1	22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
15727311	Female	35	65000	0	20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
15570769	Female	26	80000	1	17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
15606274	Female	26	52000	0	11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
15746139	Male	20	86000	1	14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
15704987	Male	32	18000	0	18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
15628972	Male	18	82000	0	22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
15697686	Male	29	80000	0	24.23155922	988.796875	19.74790765	318.3214111	0.329656571
15733883	Male	47	25000	1					

Figure A: CLASSIFICATION

Figure B: REGRESSION

Tendremos una variables especial que será nuestra variable: **TARGET (o variable dependiente)**  
Dependiendo del tipo de problema, será la naturaleza de esa variable.

En esta clase revisaremos el problema de regresión.

Mediante variables independientes, intentaremos calcular el valor de la variable dependiente.

Aplicaciones como:

- » Predicción de valor divisas.
- » Predicción del clima.
- » Estimación de valores faltantes.
- » Estimación de nivel de alguna proteína en la sangre.

Todo esto siempre será en base a otras variables que nos entregaran información para realizar la estimación o predicción.

El problema  
de regresión.

# Regresión

El problema de regresión se basa, de manera simplista, en encontrar la relación que tienen una o varias variables con otra.

Tendremos la o las variables independientes como  $X$ .

Tendremos la o las variables dependientes como  $Y$ .

En este curso, nos centraremos en el escenario donde tendremos 1 o muchas variables  $X$  y solo una variable  $Y$ .

# Regresión

Matemáticamente hablando, tendremos:

Y: Variable dependiente

$X = (X_1, X_2, \dots, X_D)$ : Las D variables independientes

$$Y = f(X) + \varepsilon$$

Puedo modelar la variable Y mediante la función f() más un término de error (que en la práctica se omite).

La idea es encontrar la función f()

# Regresión

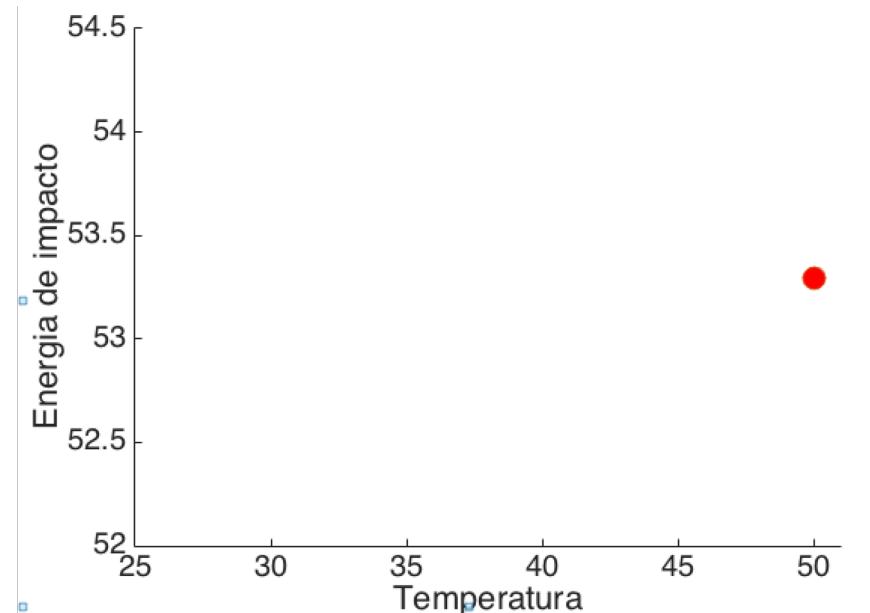
## Un ejemplo:

*“La fragilidad del acero (medida como la energía de un golpe necesaria para romperlos) depende de la temperatura”*

Medimos la fragilidad(F) de una pieza de acero y medimos la temperatura de dicha pieza (T):

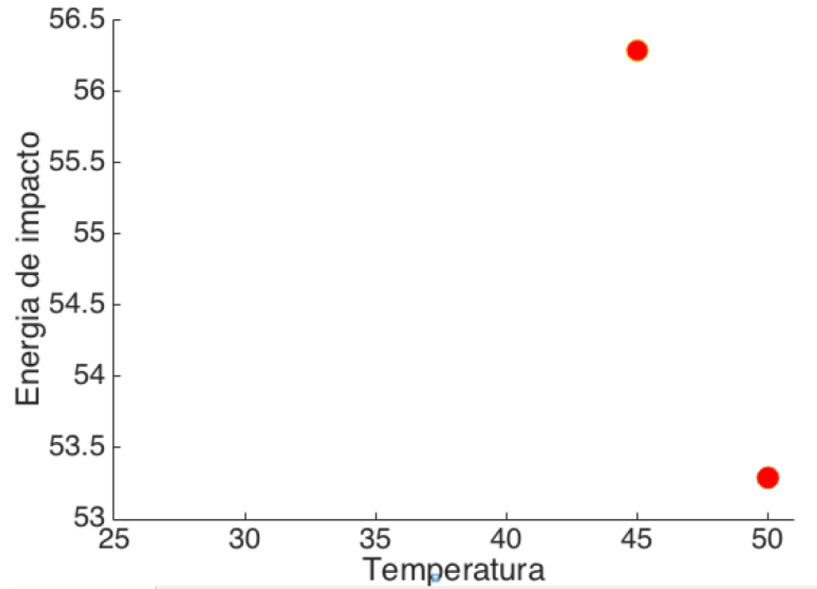
T: 50°

F: 53.2

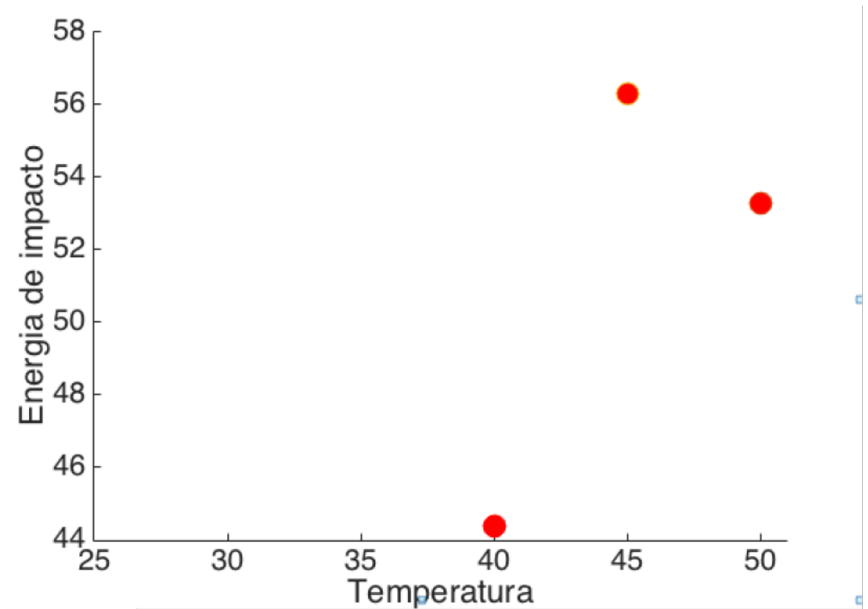


...e indicamos la medición con un punto de coordenadas  $x=50$ ,  $y=53.2$ .

# Regresión



Agregamos 1 punto



Agregamos otro...

# Regresión

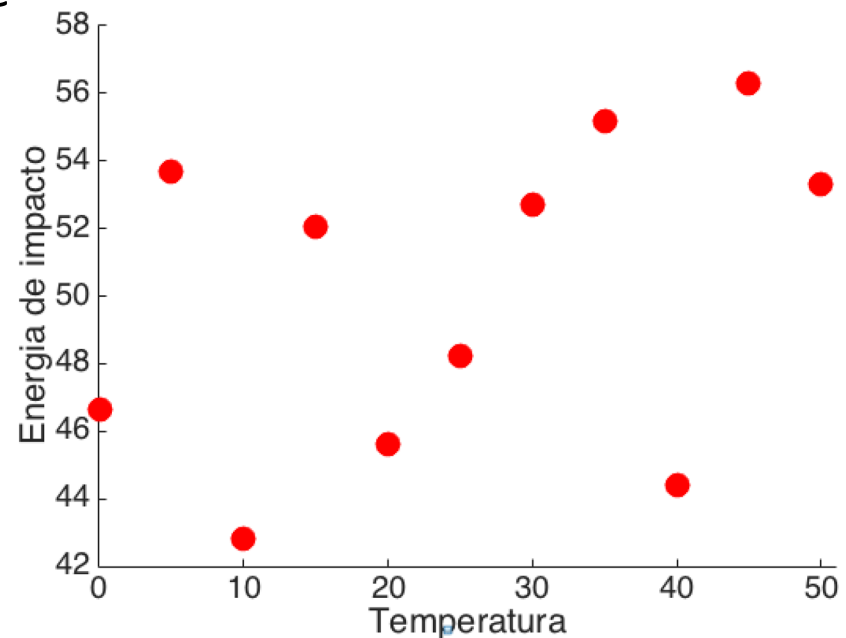
¿Vemos una relación?



Existe una forma de medir si existe relación o no.

A esto le llamaremos:

**CORRELACIÓN**





# Regresión: Correlación

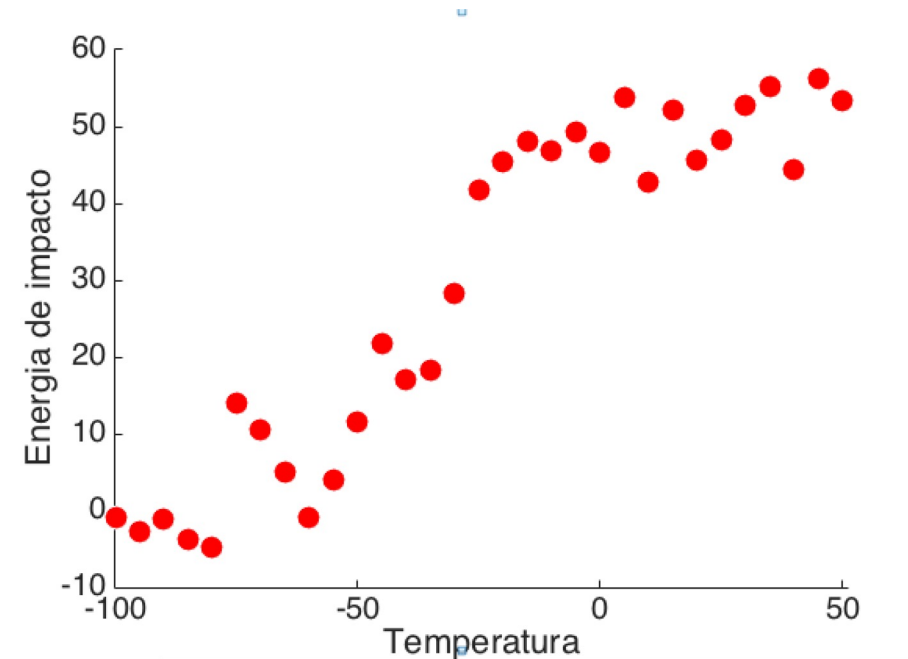
## Correlación lineal o de Pearson

no confundir con el:

Coeficiente de determinación (R2)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

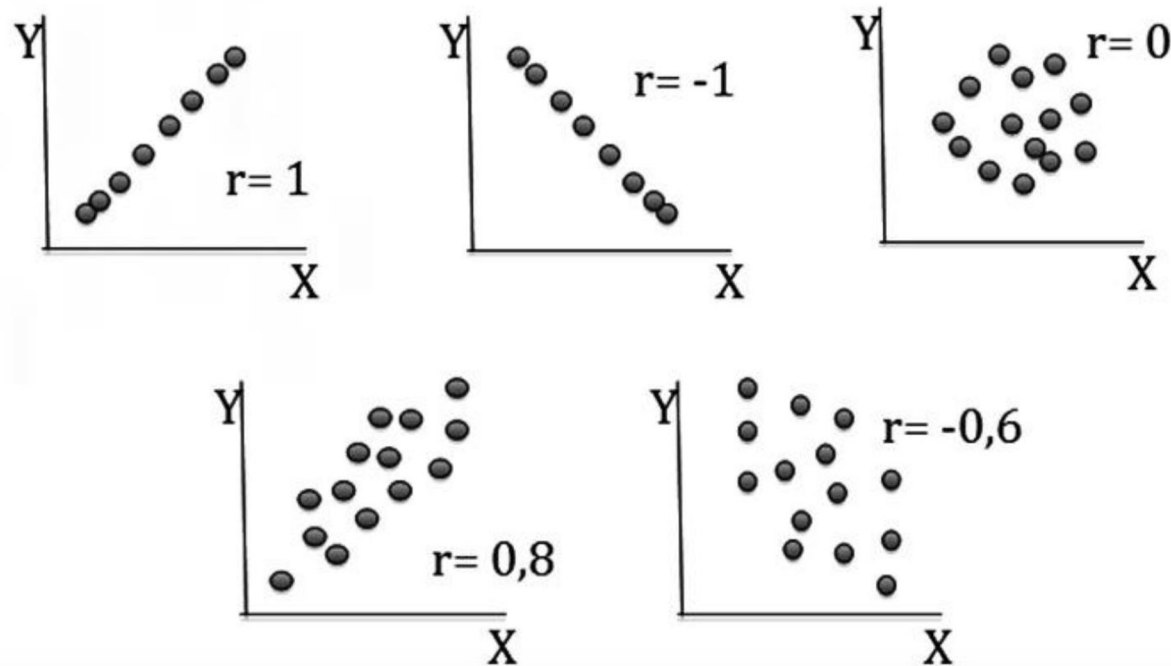
$$r = \cos(\alpha) = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$



**r: 0.93**

# Regresión

El comportamiento de la métrica  $r$ :

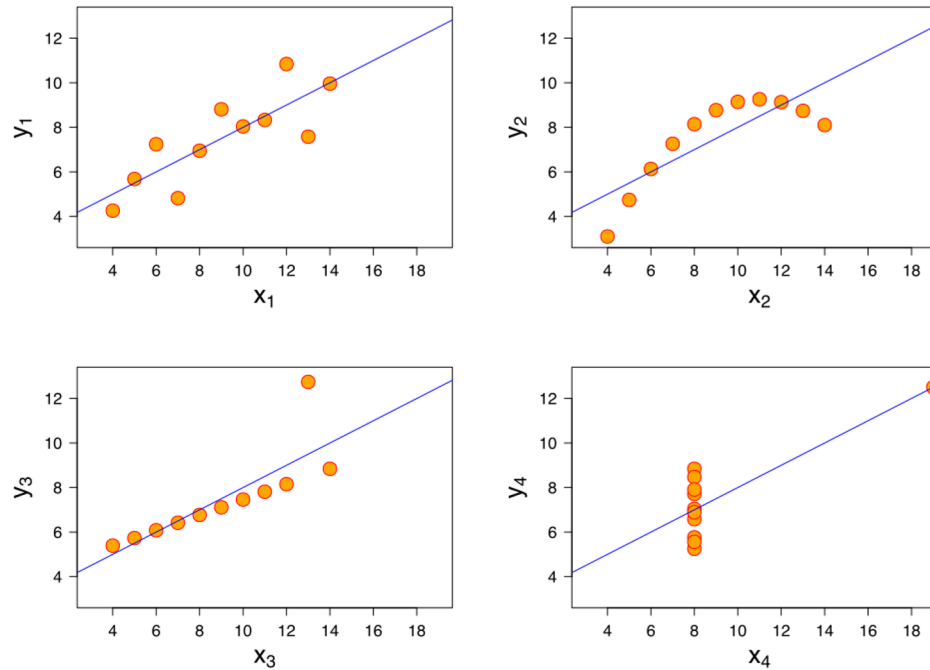


La correlación la usaremos, entre otras métricas, para medir qué tan bueno es el modelo de regresión.

# Regresión

## Cuarteto de Ascombe

(misma correlación lineal, distintas relaciones funcionales)

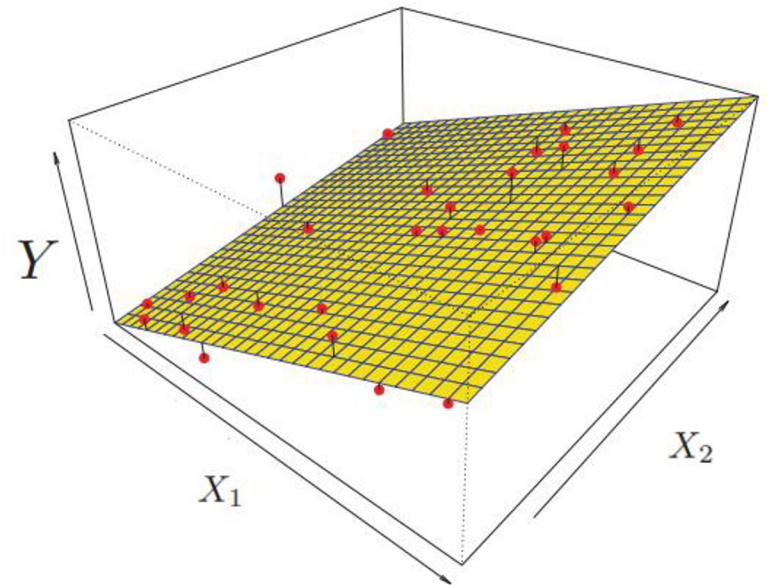
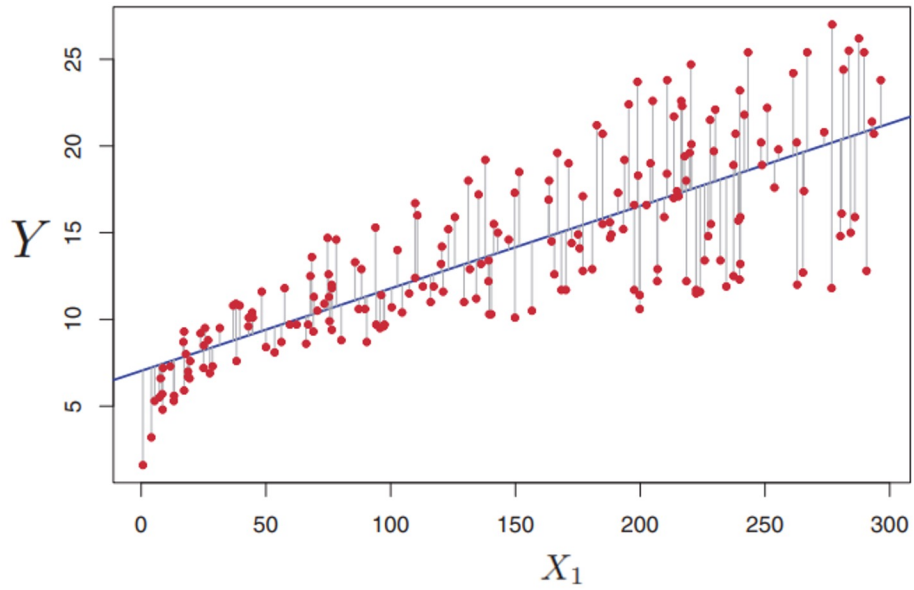


Es importante inspeccionar gráficamente (cuando se puede) los datos.

Existe una relación, pero queremos saber cuál:

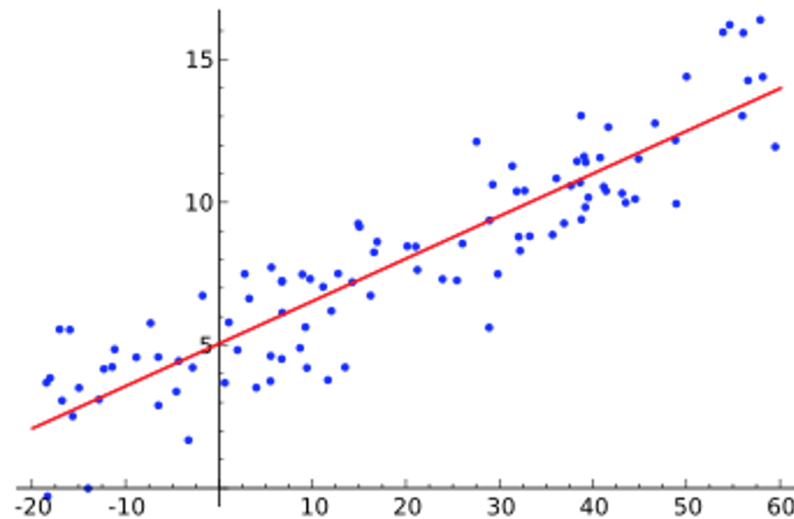
$$Y = f(X) + \epsilon$$

# Regresión



# Regresión

## La regresión lineal.



Modelo básico, pero que nos ayudará a entender otros algoritmos más sofisticados.

¿Donde estamos  
en CRISP-DM?

# CRISP-DM

