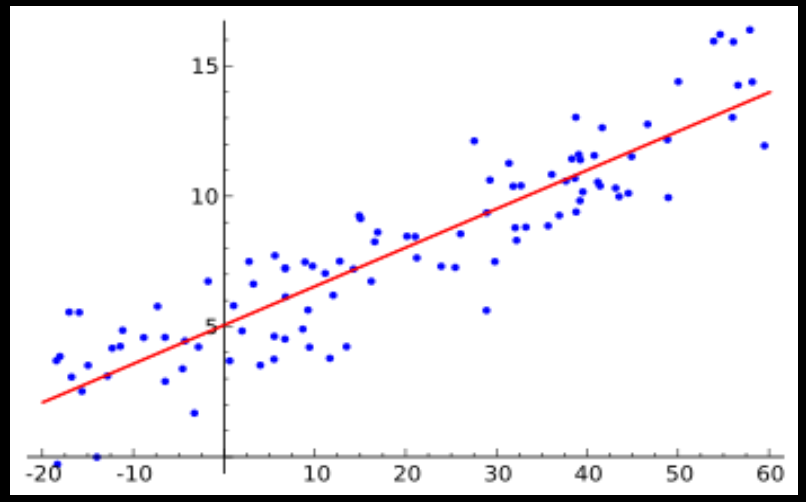
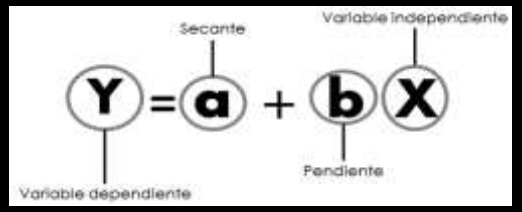
Machine Learning
Regresión Lineal

Regresión Lineal

Método estadístico que modela la relación entre una variable continua y una o más variables independientes



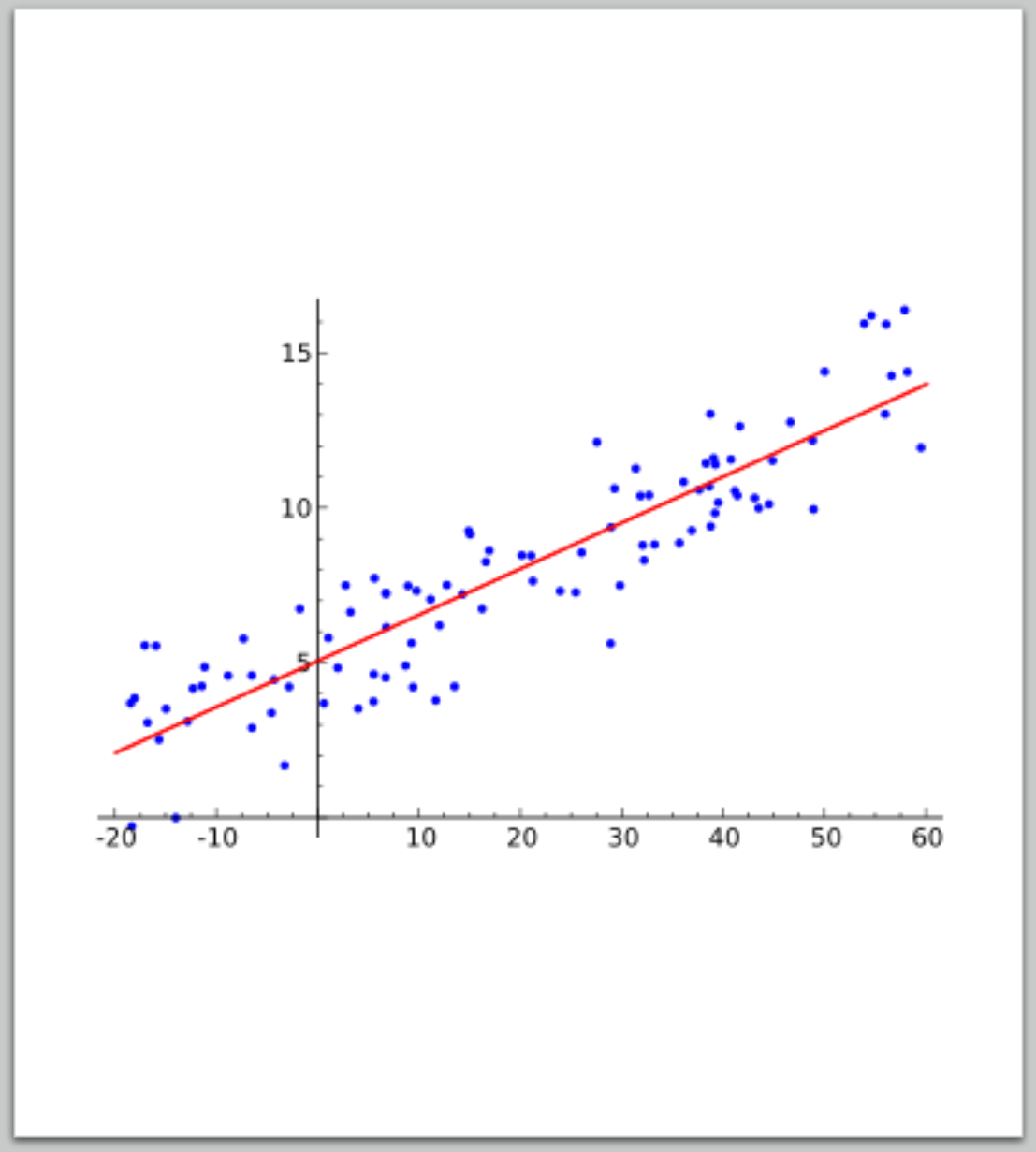
1. Estimación de las ventas a partir del gasto en marketing
2. Estimación del consumo de gasolina en función de la distancia.
3. Predicción de precios de casas en función de los metros cuadrados (entre otras variables)



Y = 5 + 6X

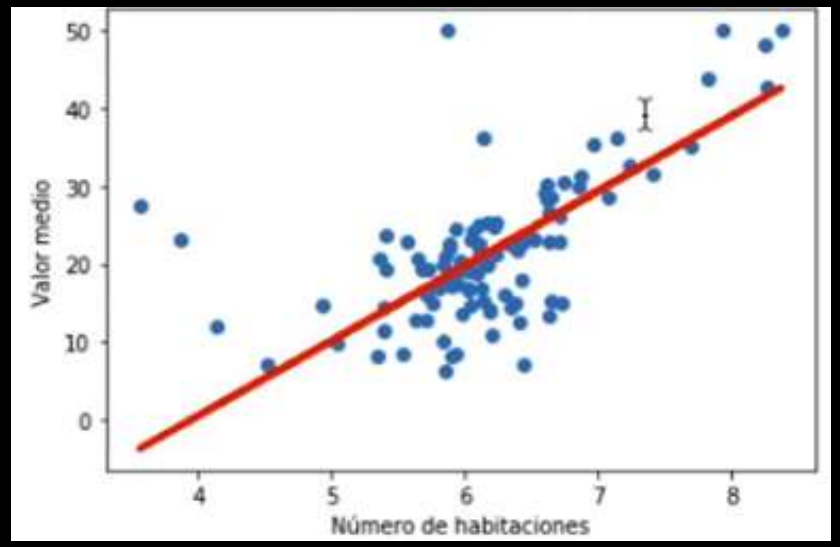
Regresión Lineal

- ¿Por qué *regresión*? Porque expresa la relación entre una variable que se llama regresando (y, dependiente) y otra que se llama regresor (x, independiente).
- ¿Por qué *lineal*? Los parámetros de la ecuación se incorporan de forma lineal
- Es una *técnica paramétrica* porque hace varias suposiciones sobre el conjunto de datos.
- Uno de los métodos estadísticos de predicción más utilizados.



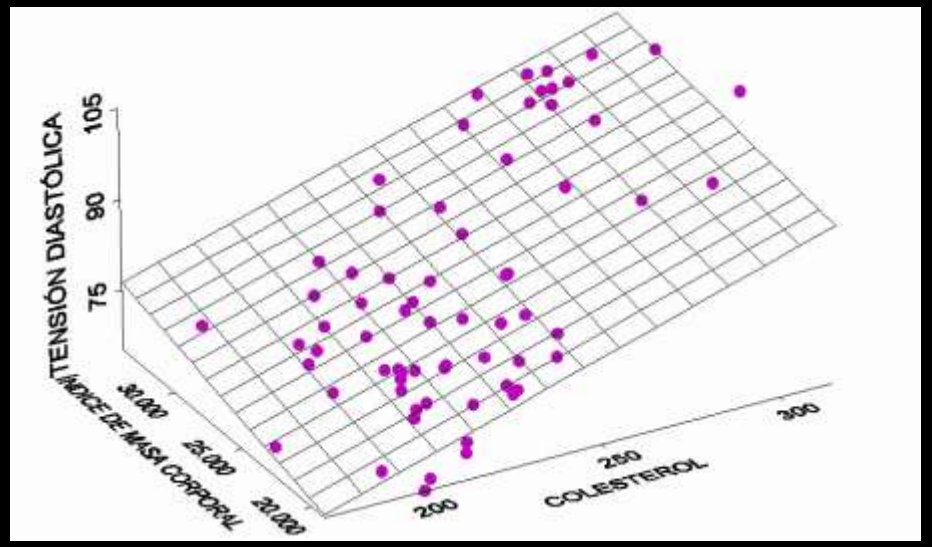
Tipos de regresión lineal

TV	radio	newspaper	sales
230.1	37.8	69.2	22100.0
44.5	39.3	45.1	10400.0
17.2	45.9	69.3	9300.0
151.5	41.3	58.5	18500.0



Regresión lineal simple

$$Y = \beta_0 + \beta_1 X_1$$



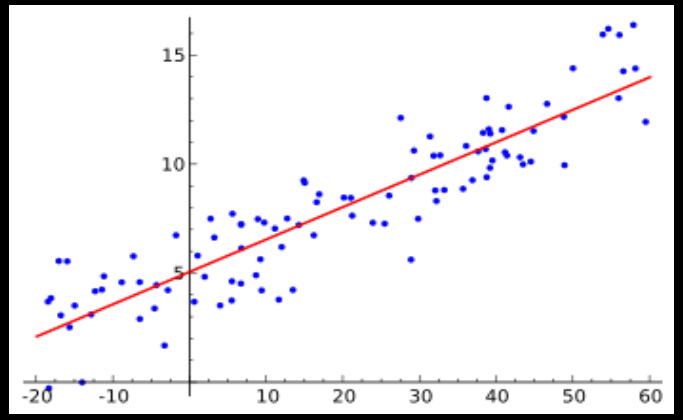
Regresión lineal múltiple

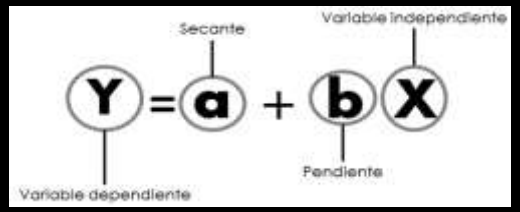
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots. + \beta_j X_j$$

¿Qué son los coeficientes?

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots. + \beta_j X_j$$

1. β_0: valor de variable respuesta para cuando todos los predictores son 0
2. β_j: cuánto aumenta la variable respuesta cuando el predictor j incrementa en una unidad





Y = 5 + 6X

¿Qué son los coeficientes?

TV	radio	newspaper	sales
230.1	37.8	69.2	22100.0
44.5	39.3	45.1	10400.0
17.2	45.9	69.3	9300.0
151.5	41.3	58.5	18500.0

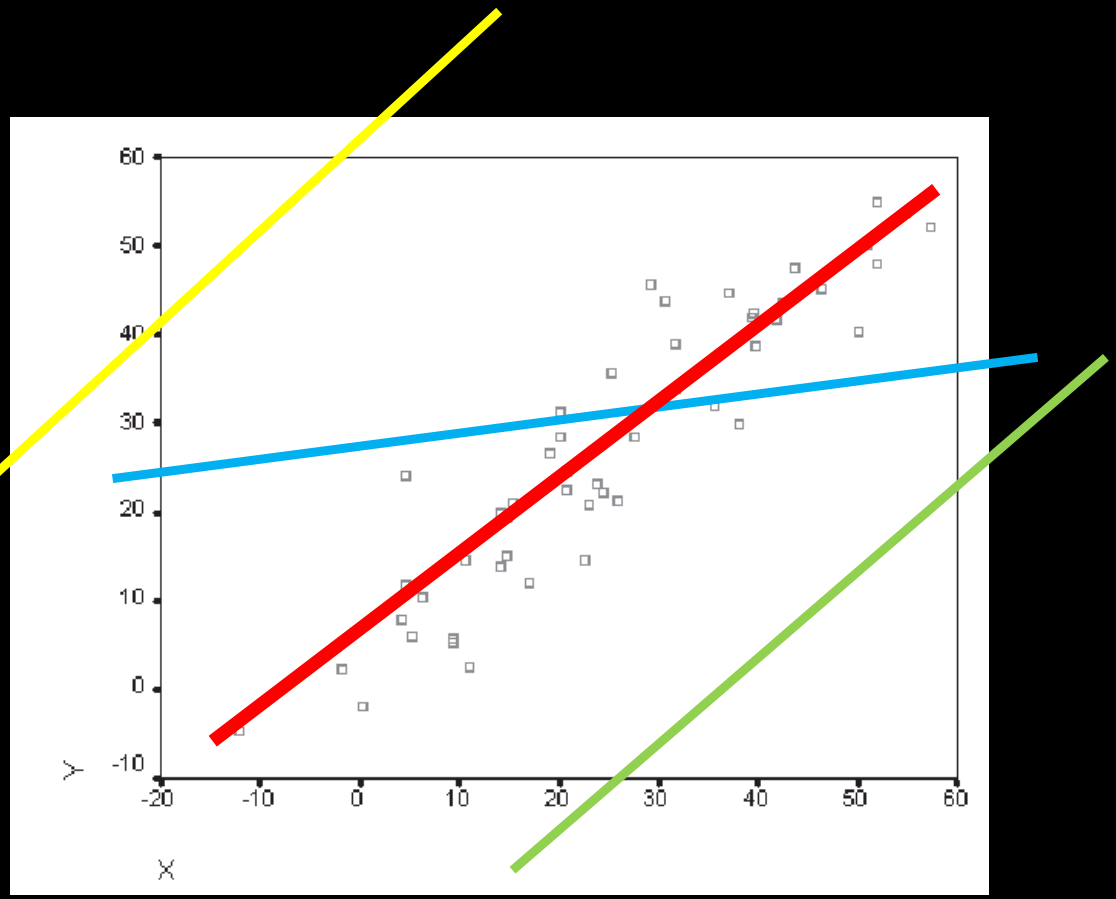
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

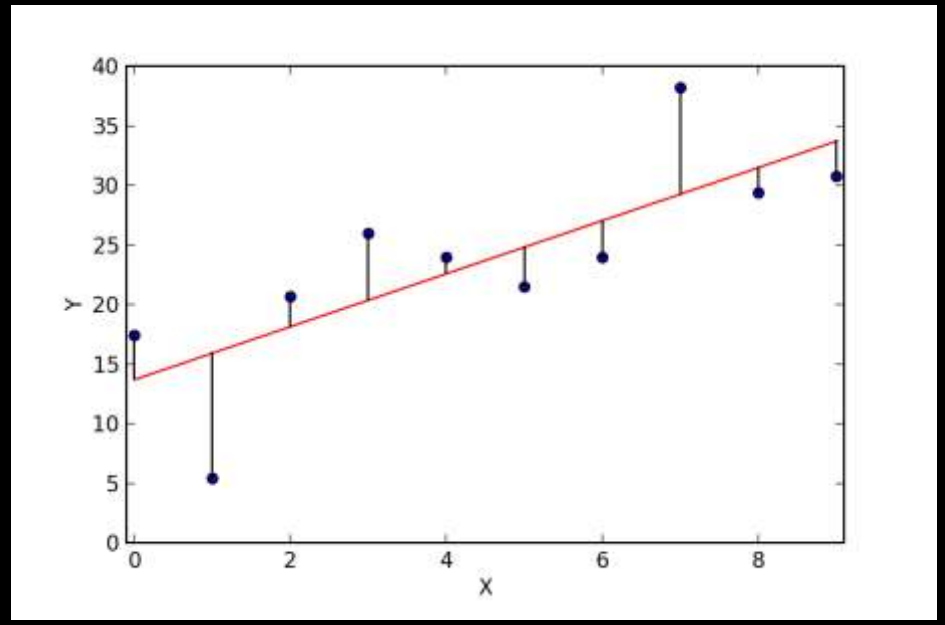
$$Sales = \beta_0 + \beta_1 *(TV) + \beta_2 (radio) + \beta_3 (newspaper)$$

$$Sales = 500 + 50*(TV) + 20(radio) + 15(newspaper)$$

¿Cómo calculo los coeficientes β?

Las β van a definir mi recta o hiperplano





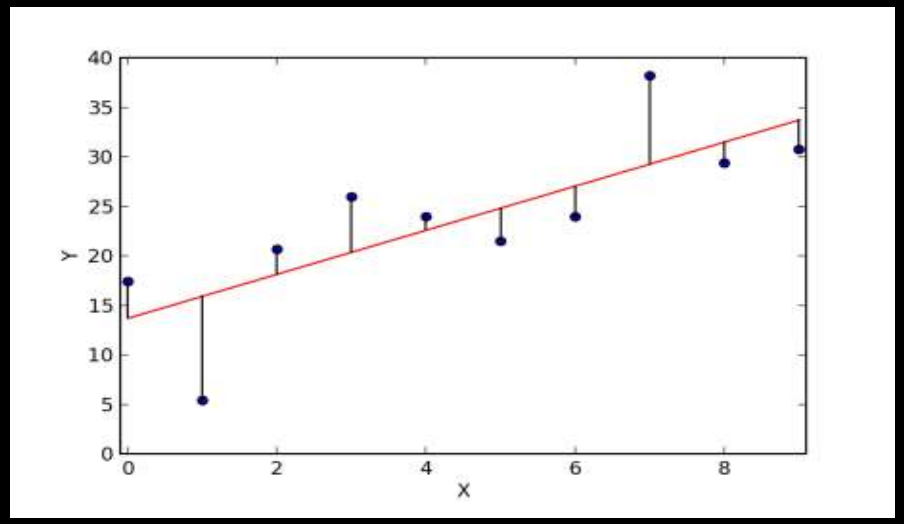
¿Cómo medimos el error cometido?

Loss function

Necesitamos una métrica que nos diga cómo de bien o mal predice el modelo:

Error cuadrático medio (Mean Squared Error)

$$MSE = \frac{1}{n}\Sigma\underbrace{\left(y - \hat{y}\right)^2}_{\text{The square of the difference between actual and predicted}}$$

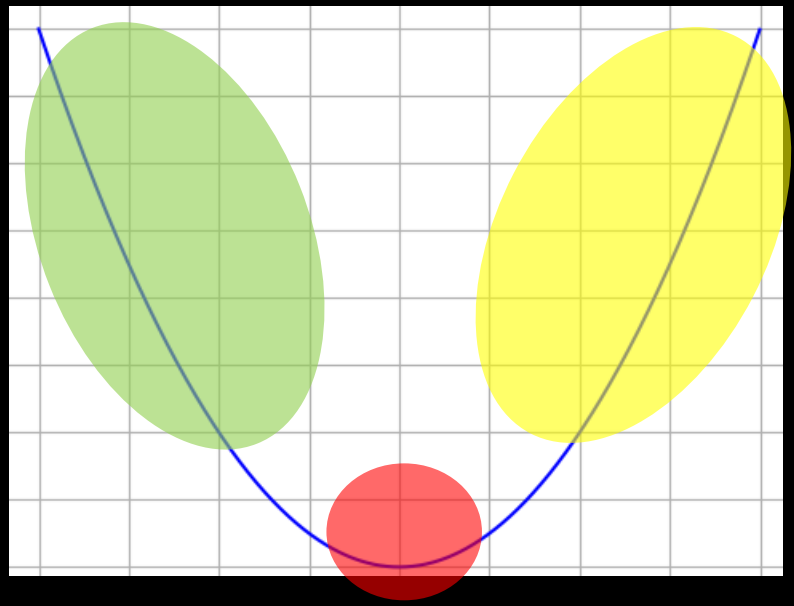


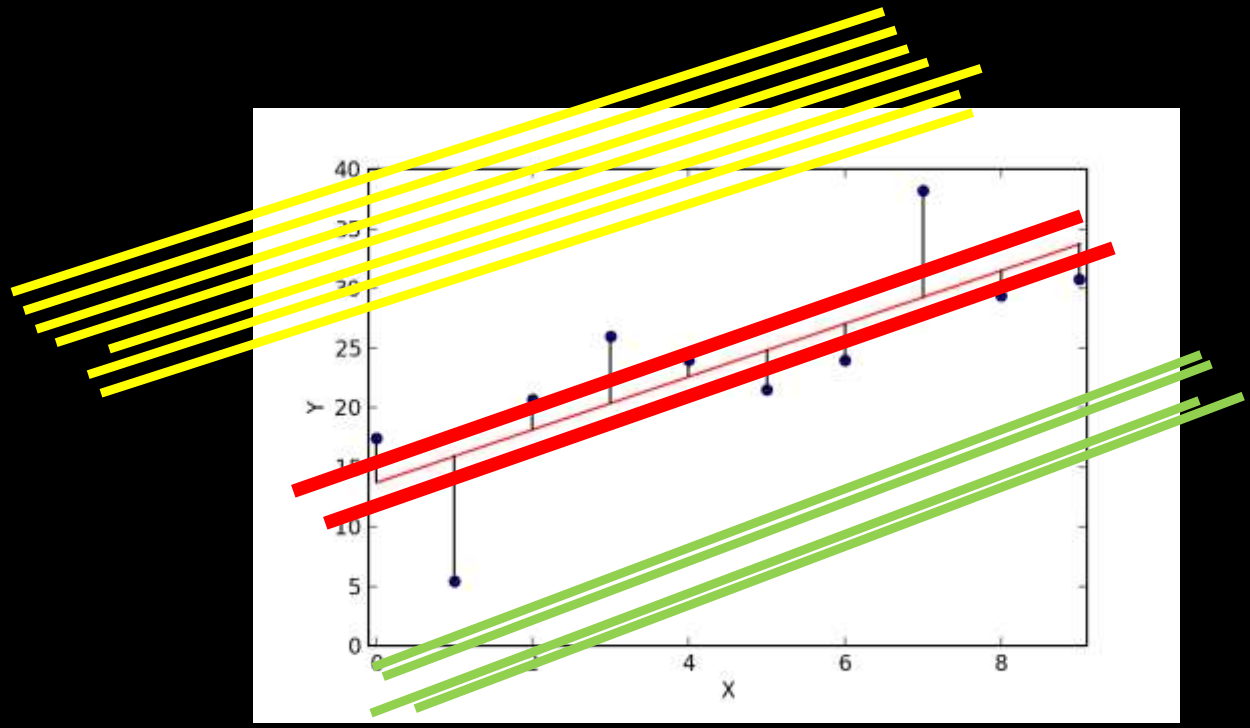
Que sería la **loss function** o función de costes de la regresión lineal

Perfecto, definida nuestra métrica de calidad del modelo, ¿ahora qué viene?
Que nuestra regresión tenga la mínima cantidad de errores, ¿cómo lo hago?
Hay que encontrar aquellos Ws que me minimicen la función de costes

Buscamos la recta que me minimice el error

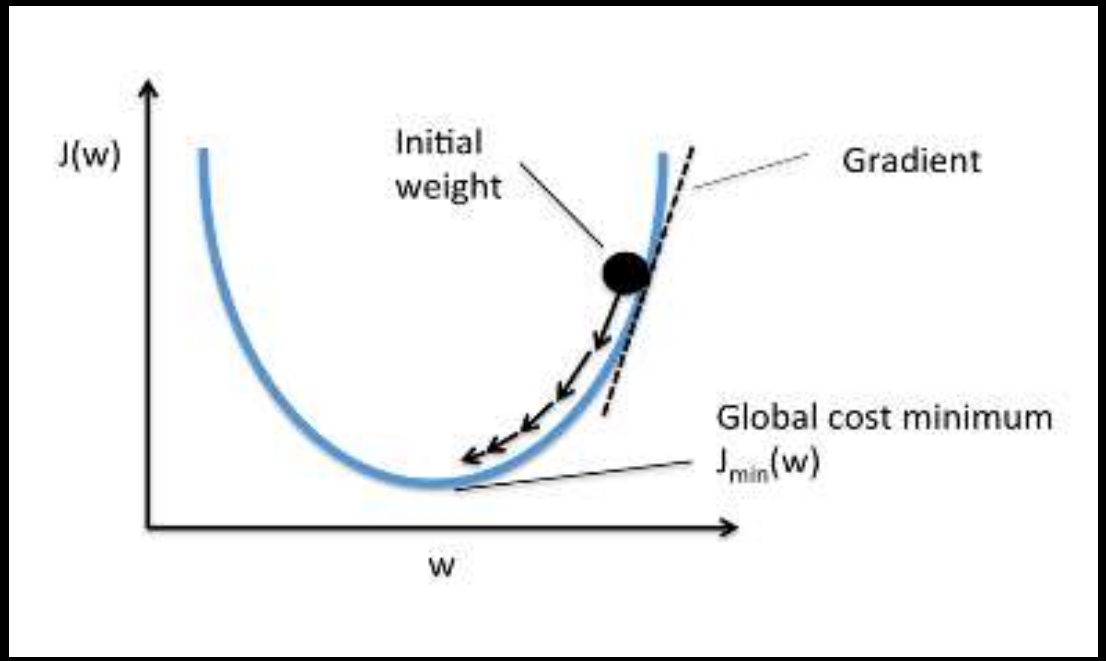
$$Y = \beta_0 + \beta_1 X_1$$





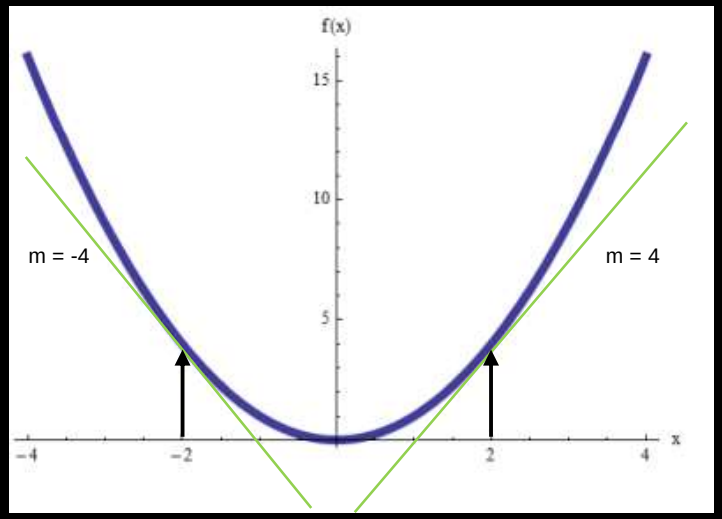
Gradient Descent

Problema de optimización matemática. El Gradient Descent es uno de los métodos más utilizados en algoritmos de aprendizaje supervisado.
¿Cuáles son los pesos W, que dan mejores resultados? Los que minimizan la función de coste



Demo

Gradient Descent



$f(x) = x^2$
$d/dx = 2x$

Hypothesis:	$h_\theta(x) = \theta_0 + \theta_1 x$
Parameters:	θ_0, θ_1
Cost Function:	$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$
Goal:	$\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

Buscamos que la pendiente sea 0
Si es positiva, estamos en el lado derecho de la curva -> Descendemos X
Si es negativa, estamos en el lado izquierdo de la curva -> Aumentamos X

Gradient Descent

1. Función de costes

$$J(w,b)=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-(wx_i+b)\right)^2$$

2. Gradiente – Derivadas parciales

$$\nabla J(w,b)=\begin{bmatrix}\frac{\delta J}{\delta w}\\ \frac{\delta J}{\delta b}\end{bmatrix}=\begin{bmatrix}\frac{2}{n}\sum_{i=1}^{n}-x_i\left(y_i-(wx_i+b)\right)\\ \frac{2}{n}\sum_{i=1}^{n}-\left(y_i-(wx_i+b)\right)\end{bmatrix}$$

3. Vemos cuánto de lejos estamos del mínimo

4. Actualizamos w y b para la siguiente iteración

$$w=w-\propto\frac{\delta J(w,b)}{\delta w}$$

$$b=b-\propto\frac{\delta J(w,b)}{\delta b}$$

Learning rate (α): parámetro que determina el salto, definido por nosotros.

5. Acaba el algoritmo cuando alcanzamos la convergencia

¿Cómo interpreto los pesos (w)?

Coeficientes

Si queremos predecir el precio de casas de un DF, podríamos obtener los siguientes coeficientes:

	Coefficient
Avg. Area Income	21.625799
Avg. Area House Age	165590.392746
Avg. Area Number of Rooms	119827.783390
Avg. Area Number of Bedrooms	2361.095262
Area Population	15.216581

E interpretaríamos la regresión lineal como:

y = w1*x1 + w2*x2 + w3*x3 + w4*x4 + w5*x5

Precio casas = 21.6 * (Avg. Area Income) + 165590.4 * (Avg. Area House Age)…..

¿Cómo se interpreta esto? Por cada unidad de *Avg. Area Income*, aumenta 21.6 el precio

Feature importance

Vale, entonces cuanto más alto es el coeficiente, mayor es la importancia de la variable...

	Coefficient
Avg. Area Income	21.625799
Avg. Area House Age	165590.392746
Avg. Area Number of Rooms	119827.783390
Avg. Area Number of Bedrooms	2361.095262
Area Population	15.216581

	coefficient
Avg. Area House Age	165590.392746
Avg. Area Number of Rooms	119827.783390
Avg. Area Number of Bedrooms	2361.095262
Avg. Area Income	21.625799
Area Population	15.216581

NO! Estamos comparando unidades diferentes. ¿La edad de la casa es más importante que la media de ingresos de la zona?

Precio ($) = Edad(años) + Habitaciones(nº habitaciones)…

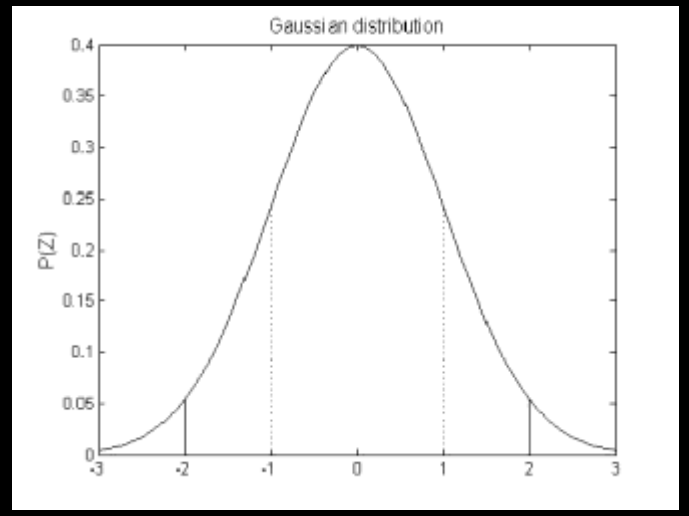
¿Solución? Estandarizar los datos

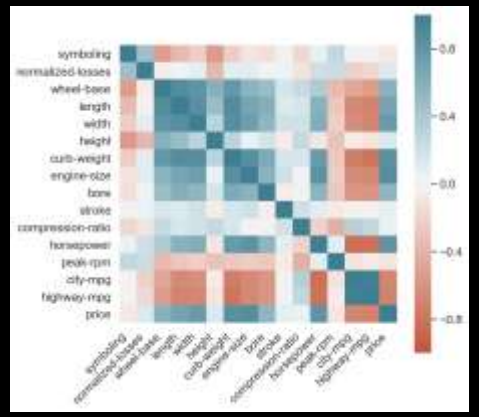
$$z = \frac{x - \mu}{\sigma}$$

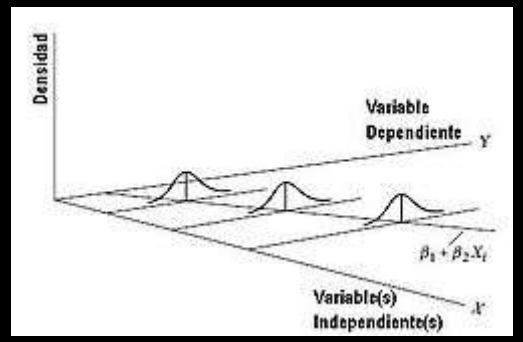
Condiciones para la regresión lineal

Condiciones I

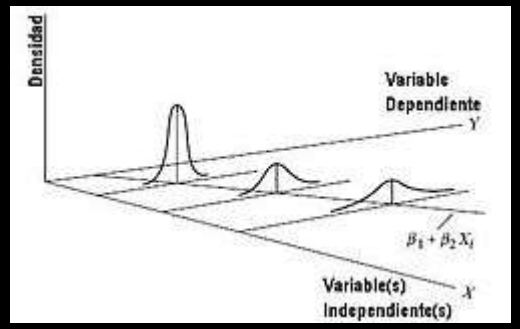
1. Distribución normal del target
2. No colinealidad o multicolinealidad. Correlación entre los predictores. Lo solucionamos eliminando uno de ellos. Hay que garantizar la independencia entre todos ellos. Un coeficiente representa el cambio en la variable dependiente cuando la variable independiente se modifica en uno su unidad, manteniendo el resto de los coeficientes constantes.
3. Relación lineal entre target y predictores. Matriz de correlación
4. Homocedasticidad. Varianza constante de los errores a lo largo de las observaciones.







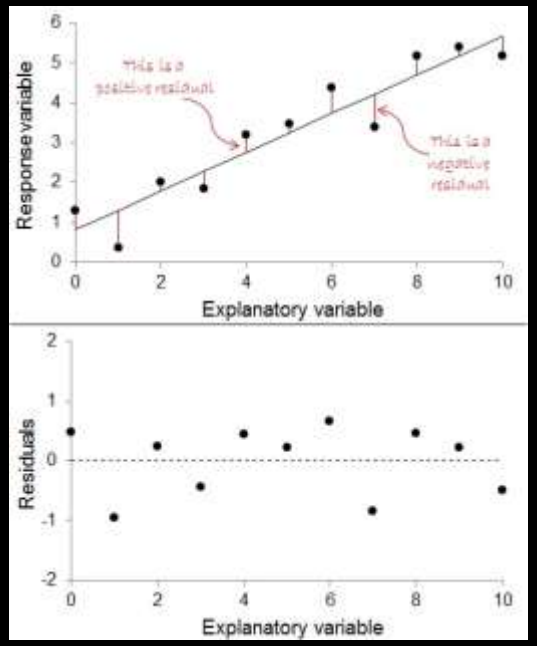
Distribución Homocedástica

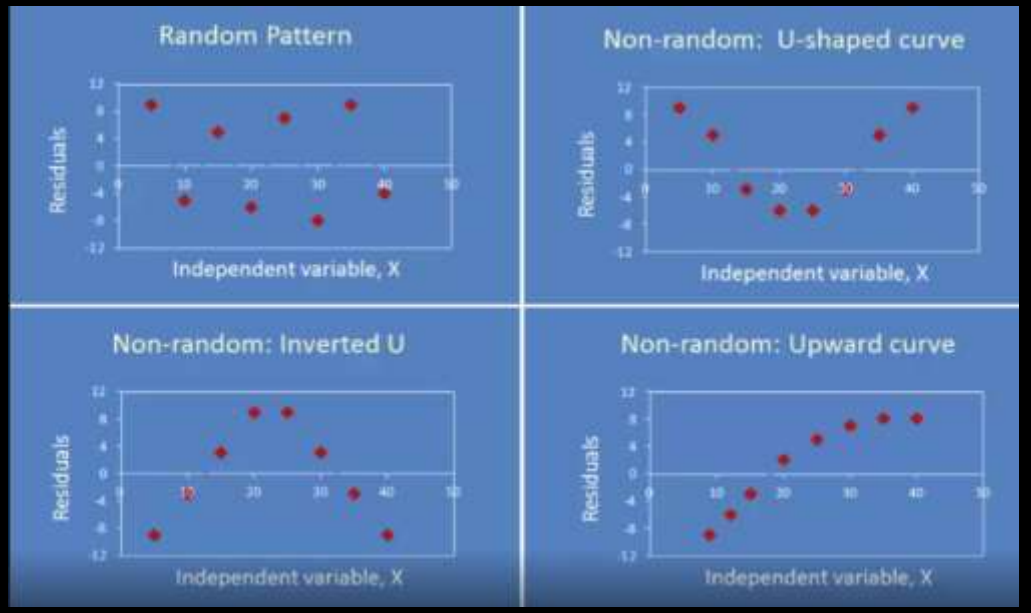


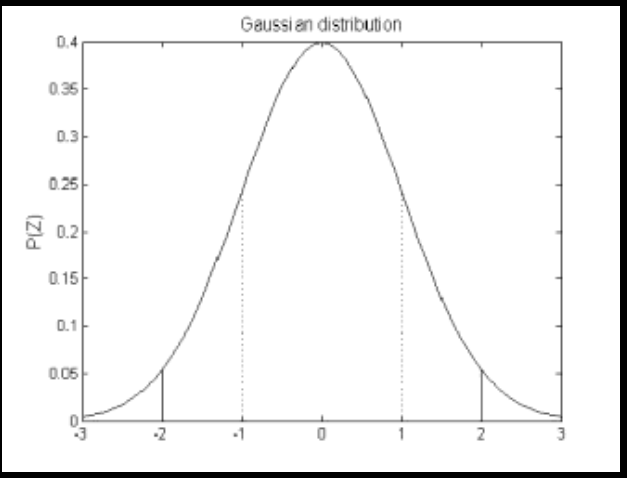
Distribución Heterocedástica

Condiciones II

1. Independencia de los residuos (errores)
2. Normalidad de los residuos (errores)
3. Valores atípicos o influyentes, outliers, solo valores cuantitativos…







R-Squared

Coeficiente de determinación. Mide cuánto de bien una regresión se ajusta a los datos. También se define como la porción de variación de la variable dependiente (y) predecible mediante la independiente (x). Va de [0,1]. Cuanto mejor se ajuste, más se acercará a 1. Cuanto más cercano a 0, menos fiable será.

How to know if the model is best fit for your data?

The most common metrics to look at while selecting the model are:

STATISTIC	CRITERION
R-Squared	Higher the better *(> 0.70)*
Adj R-Squared	Higher the better
F-Statistic	Higher the better
Std. Error	Closer to zero the better
t-statistic	Should be greater 1.96 for p-value to be less than 0.05
AIC	Lower the better
BIC	Lower the better
Mallows cp	Should be close to the number of predictors in model
MAPE (Mean absolute percentage error)	Lower the better
MSE (Mean squared error)	Lower the better
Min_Max Accuracy => mean(min(actual, predicted)/max(actual, predicted))	Higher the better

Evaluación del modelo

Regresión lineal: errores

- Mean Absolute Error (MAE). Errores en unidades del target

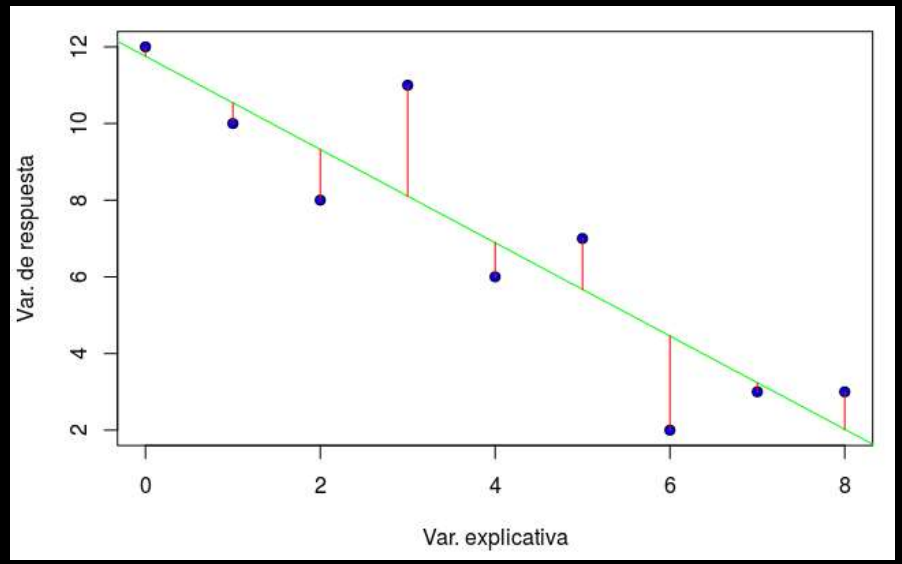
$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

- Mean Squared Error (MSE). No se interpreta bien al ir al cuadrado, pero enfatiza mucho más los errores altos

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

- Root Mean Squared Error (RMSE)

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}}.$$



Bibliografía

https://learning.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

https://www.cienciadedatos.net/documentos/py10-regresion-lineal-python.html

https://machinelearningmastery.com/gradient-descent-for-machine-learning/

https://towardsdatascience.com/gradient-descent-animation-1-simple-linear-regression-e49315b24672