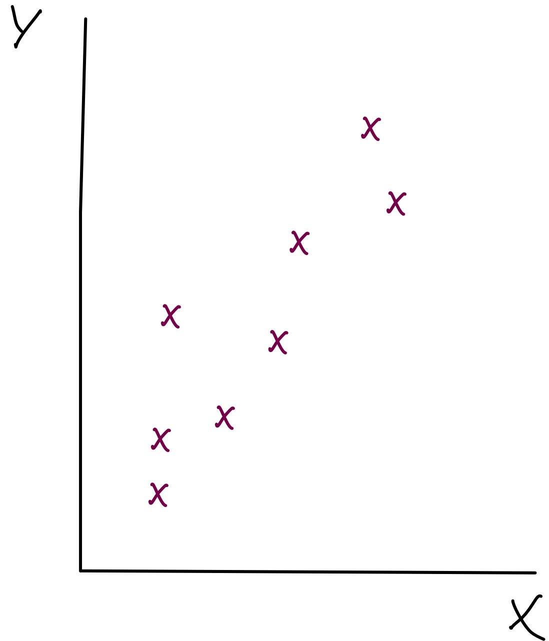
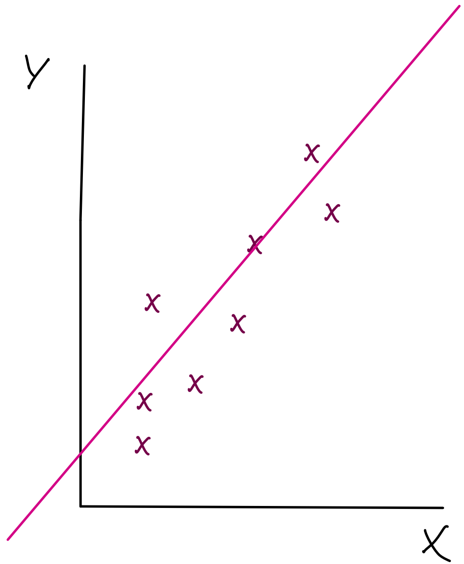
Hoy:

- [ ] Regresión lineal
- [ ] Mínimos cuadrados (RSS)
- [ ] Decenso de gradiente
- [ ] Coeficiente de determinación
- [ ] Práctica linear regressor de Sklearn

Regresión lineal



Regresión lineal



Consiste en determinar un modelo lineal para describir o aproximar los datos

Consideremos un conjunto de datos con p atributos y una variable objetivo que denotaremos Y.

Usaremos $X_1, X_2, \ldots, X_p$ para denotar los p atributos

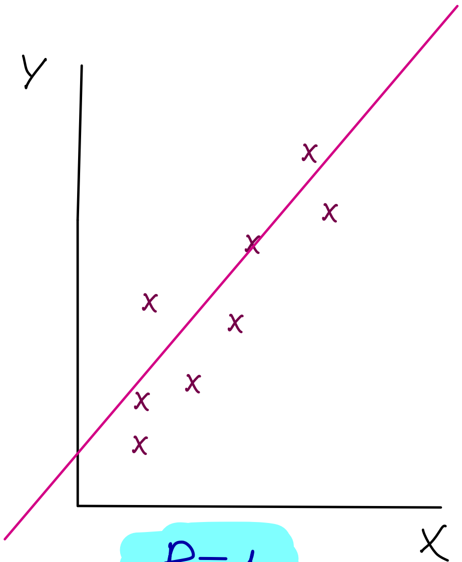
y $X = (X_1, X_2, \ldots, X_p)^T$

El modelo tendrá la forma

$$\boxed{f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j}$$

parámetros

Regresión lineal



$p = 1$

Consiste en determinar un modelo lineal para describir o aproximar los datos

Consideremos un conjunto de datos con p atributos y una variable objetivo que denotaremos Y.

Usaremos $X_1, X_2, \ldots, X_p$ para denotar los p atributos y $X = (X_1, X_2, \ldots, X_p)^T$

El modelo tendrá la forma

$$\boxed{f(X) = \underbrace{\beta_0}_{\text{parámetros}} + \sum_{j=1}^{p} X_j \beta_j}$$

parámetros

Observación. En ocasiones se designa $X_0 = \mathbb{1}$ → vector de 1's

y se expresa $f(X)$ con un sólo término

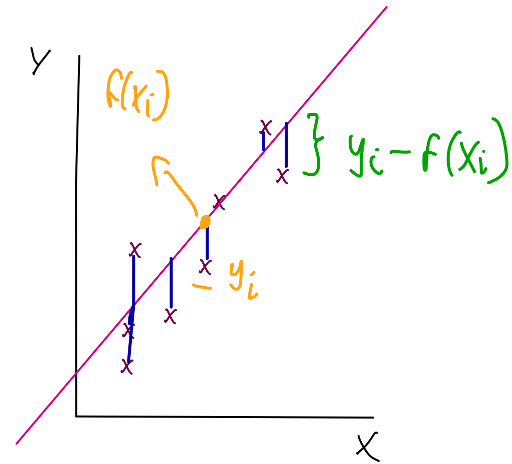
$$f(X) = \sum_{j=0}^{p} X_j \beta_j$$

$$\boxed{f(X) = \underbrace{\beta_0}_{\text{parámetros}} + \sum_{j=1}^{p} X_j \underbrace{\beta_j}_{\text{parámetros}}}$$

¿Cómo estimar los parámetros β_i?

Dado un conjunto de datos

$$D = (\vec{x}_1, y_1) \dots (\vec{x}_N, y_N)$$



Se quisiera minimizar la

Suma Residual de Cuadrados

$$\boxed{RSS(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2}$$

$$= \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{i=1}^{p} x_{ij} \beta_j \right)$$

$$\boxed{RSS(\beta) = \sum_{i=1}^{N} \left(y_i - f(x_i)\right)^2}$$

Columna atributo 1

atributo N

Sea $\mathbb{X}$ la matriz

$$N \text{ registros} \left\{ \begin{bmatrix} 1 & x_{11} & & x_{1p} \\ 1 & & & \\ \vdots & x_{i1} & & x_{ip} \\ 1 & & & \\ 1 & & & \end{bmatrix} \right.$$

$$\underbrace{\qquad\qquad\qquad}_{p}$$

Tamaño

$N \times (p+1)$

y sea $\vec{y}$ el vector de salidas en el conjunto de entrenamiento

$$RSS(\beta) = \left(\vec{y} - \mathbb{X}\vec{\beta}\right)^T \left(\vec{y} - \mathbb{X}\vec{\beta}\right)$$

$$RSS(\beta) = (\vec{y} - \mathbb{X}\vec{\beta})^T (\vec{y} - \mathbb{X}\vec{\beta})$$

Esta es una función cuadrática en los parametros $\beta_0, \dots, \beta_p$.
Si derivamos con respecto a los β_i obtenemos

$$\frac{\partial RSS}{\partial \beta} = -2\,\mathbb{X}^T (\vec{y} - \mathbb{X}\vec{\beta})$$

$$\frac{\partial^2 RSS}{\partial \beta \partial \beta^T} = 2\,\mathbb{X}^T \mathbb{X}$$

Si $\mathbb{X}$ tiene rango completo entonces $\mathbb{X}^T\mathbb{X}$ es semidefinida positiva y el mínimo se encontraría igualando la primera derivada a cero

$$\mathbb{X}^T (\vec{y} - \mathbb{X}\vec{\beta}) = 0 \qquad \hat{\beta} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \vec{y}.$$

El método anterior nos provee de una solución matemática precisa para estimar el vector de parámetros $\vec{\beta}$.

Pero.... es computacionalmente costoso

¿Otra opción?

$\boxed{\text{Descenso de gradiente}}$

Un método, en este caso computacionalmente menos costoso para optimizar RSS, y en general muy utilizado en ML incluyendo las redes neuronales

Queremos minimizar:

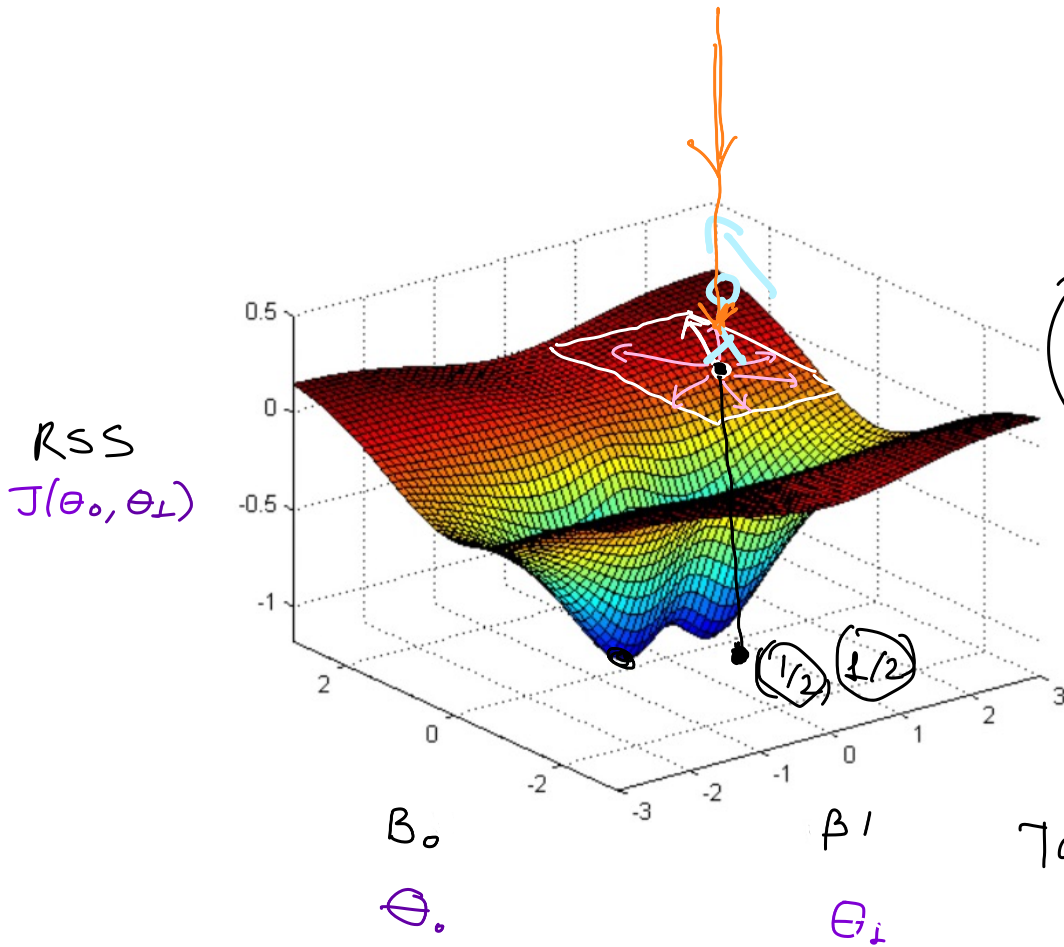
$$RSS(\beta) = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)$$

Ejemplo $p = 1$

$$RSS(\beta) = \sum_{i=1}^{N} y_i - \beta_0 - X_{i1}\beta_1$$

Dos variables!

Una función continua de dos variables se puede imaginar como una superficie



$J(\theta_0, \theta_1)$ $(3, -10)$

$\nabla J(\theta_0, \theta_1) = \left(\frac{\partial J}{\partial \theta_0}, \frac{\partial J}{\partial \theta_1}\right)$

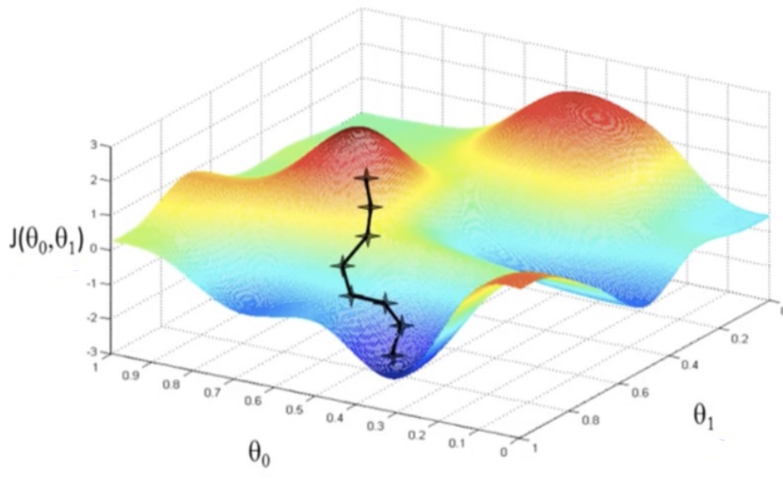
Damos un paso

en dirección opuesta

al gradiente

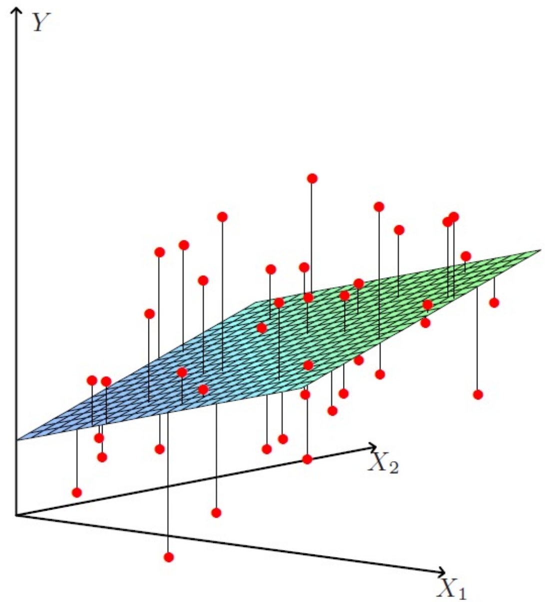
Tamaño del paso α

learning rate



J(θ0,θ1)
3
2
1
0
-1
-2
-3
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
θ0
0
0.2
0.4
0.6
0.8
1
θ1

Una vez tenemos el modelo

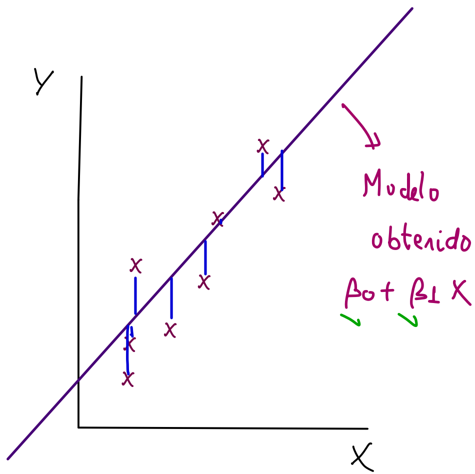


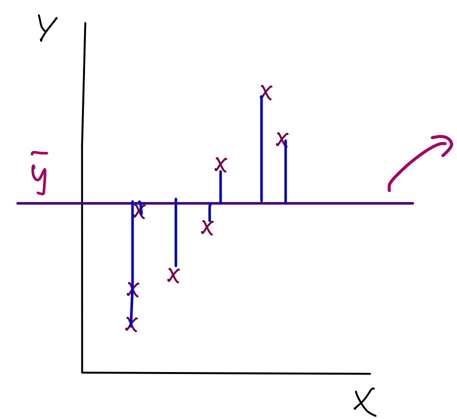
¿Cómo podemos medir qué tan bueno es?

Coeficiente de determinación R^2

Sirve para medir qué tanto la variable dependiente es explicada por la(s) variable(s) independiente(s).

¿Cómo se mide? consideremos los cuadrados de las diferencias en 2 escenarios:





$$RSS = \sum_{i=1}^{N} (y_i - f(x_i))^2$$

$$SS(mean) = \sum_{i=1}^{N} (y_i - \bar{y})^2$$

R^2 compara qué tanto disminuye el error nuestro modelo con respecto al modelo horizontal $\bar{y}$.

$$R^2 = 1 - \frac{RSS}{SS(mean)}$$

$$R^2 = \frac{SS(mean) - RSS}{SS(mean)}$$

$$0 \leqslant R^2 \leqslant 1$$