Phaw AI

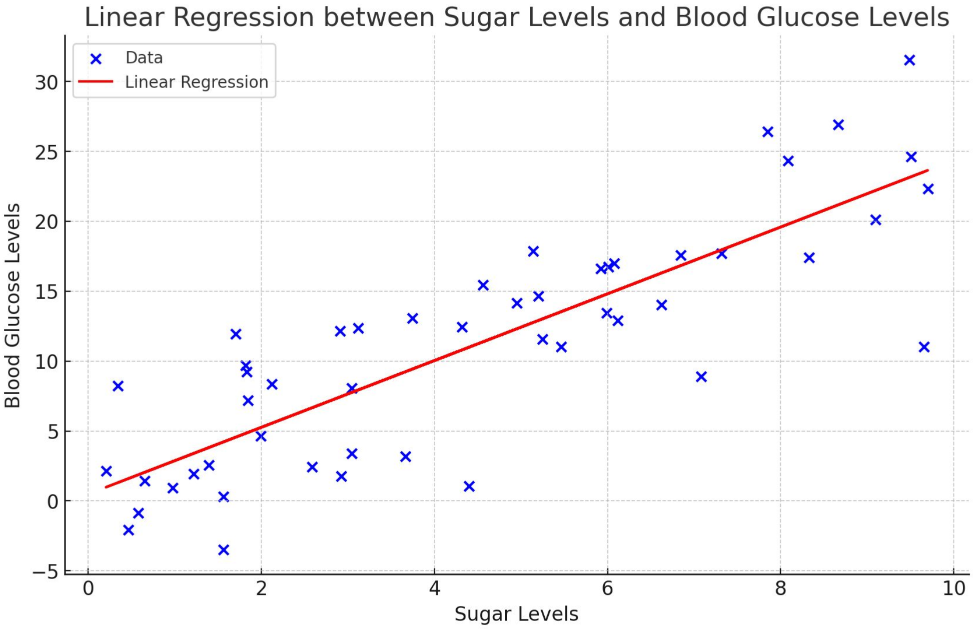
Dr. Cristian López Del Alamo
clopezd@utec.edu.pe
ULASALLE

Phaw AI

Dr. Cristian López Del Alamo
clopezd@utec.edu.pe
ULASALLE

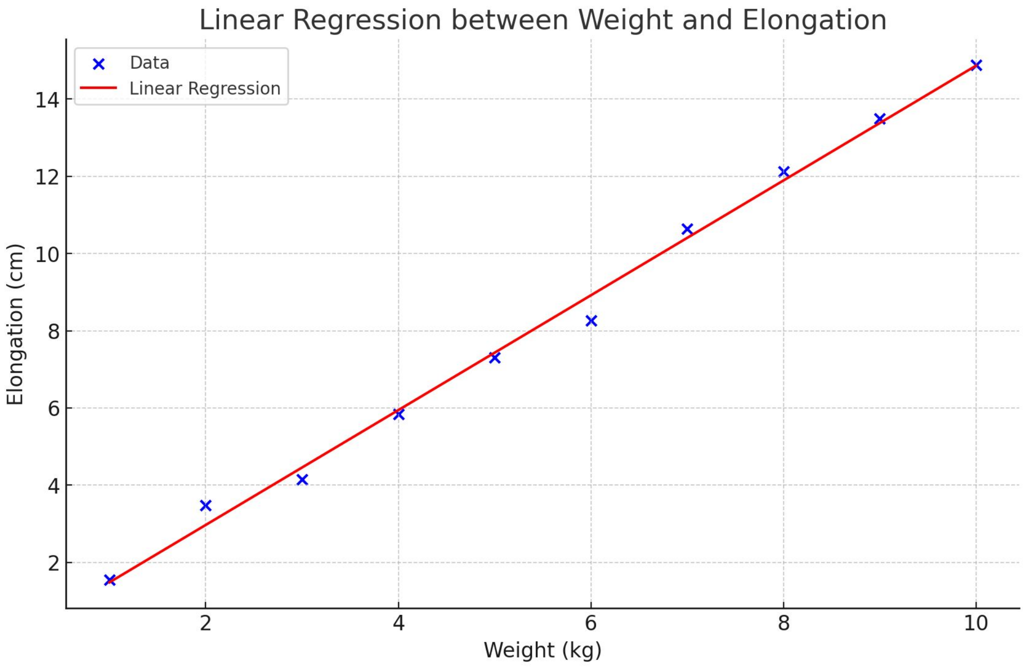
Objective: Understand the idea behind univariate linear regression

Prediction of blood sugar levels



Sugar Levels	Blood Glucose Levels
3.75	13.06
9.51	24.62
7.32	17.72
5.99	13.46
1.56	-3.49

Prediction of the elongation of a spring of static size



Weight (kg)	Elongation (cm)
1	1.55
2	3.48
3	4.15
4	5.84
5	7.30
6	8.27
7	10.65
8	12.13
9	13.50
10	14.88

Model

$$L = xw + b$$

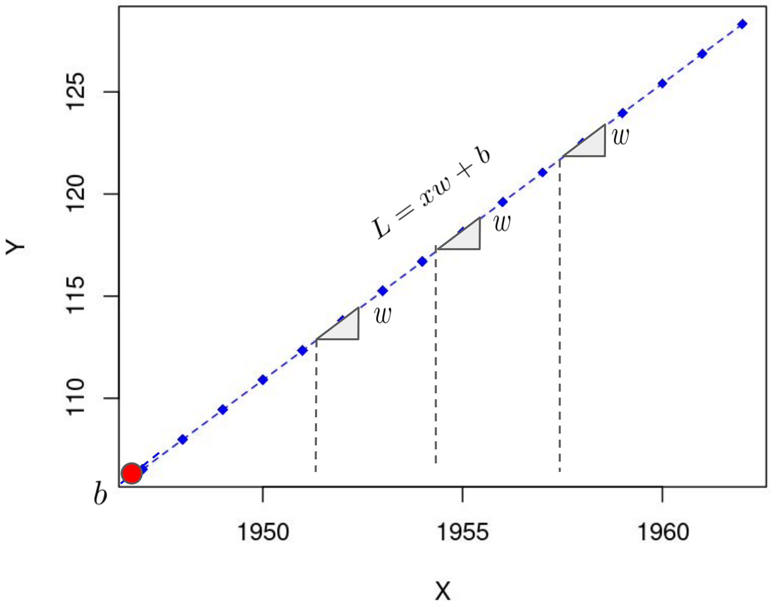
Parameters

Model

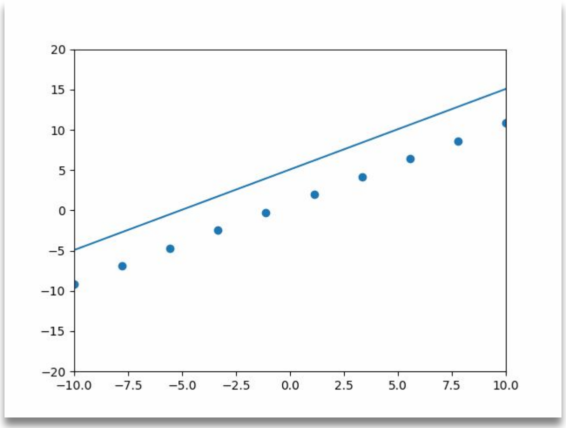
$$L = xw + b$$

bias

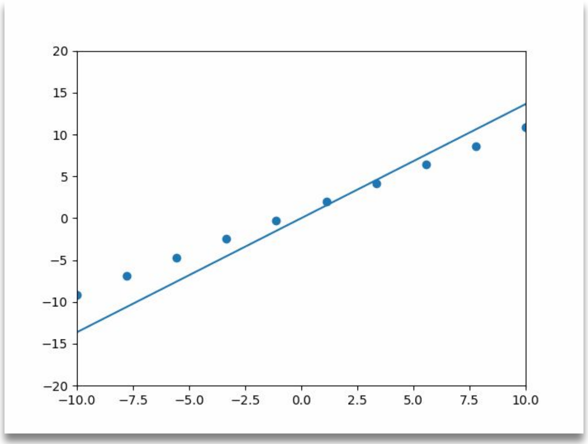
Slope



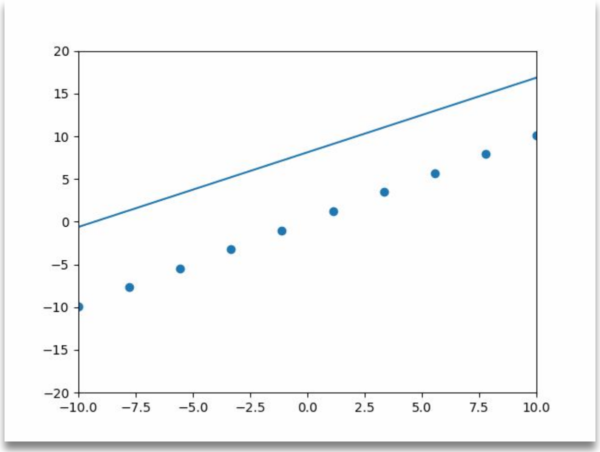
$$L = xw + \boxed{b}$$



$$L = x\boxed{w} + b$$

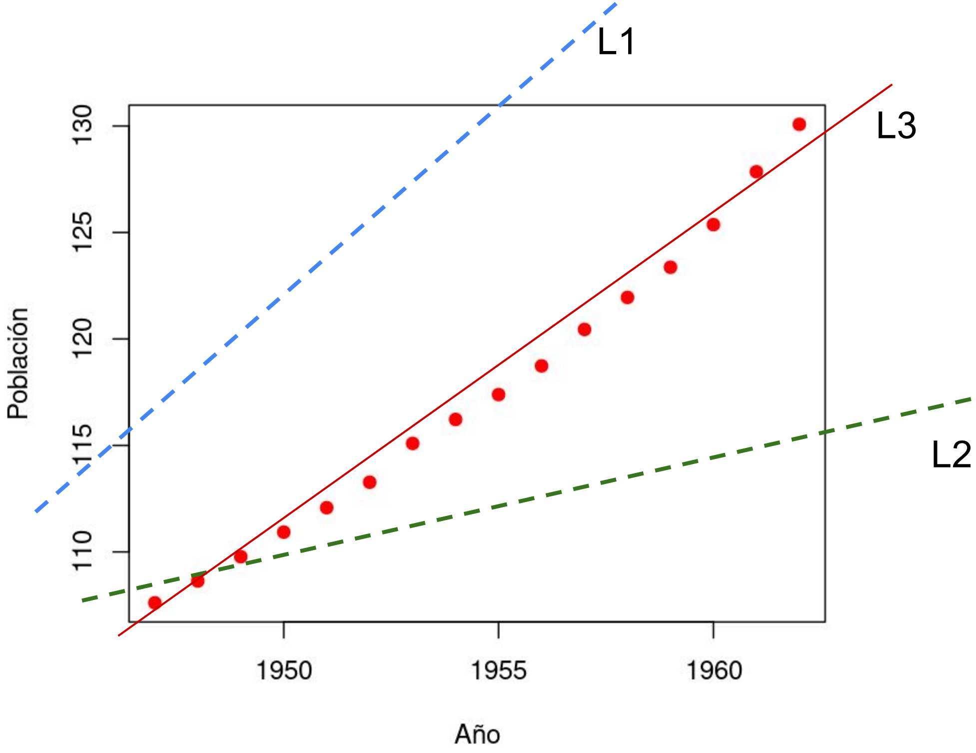


$$L = x\boxed{w} + \boxed{b}$$

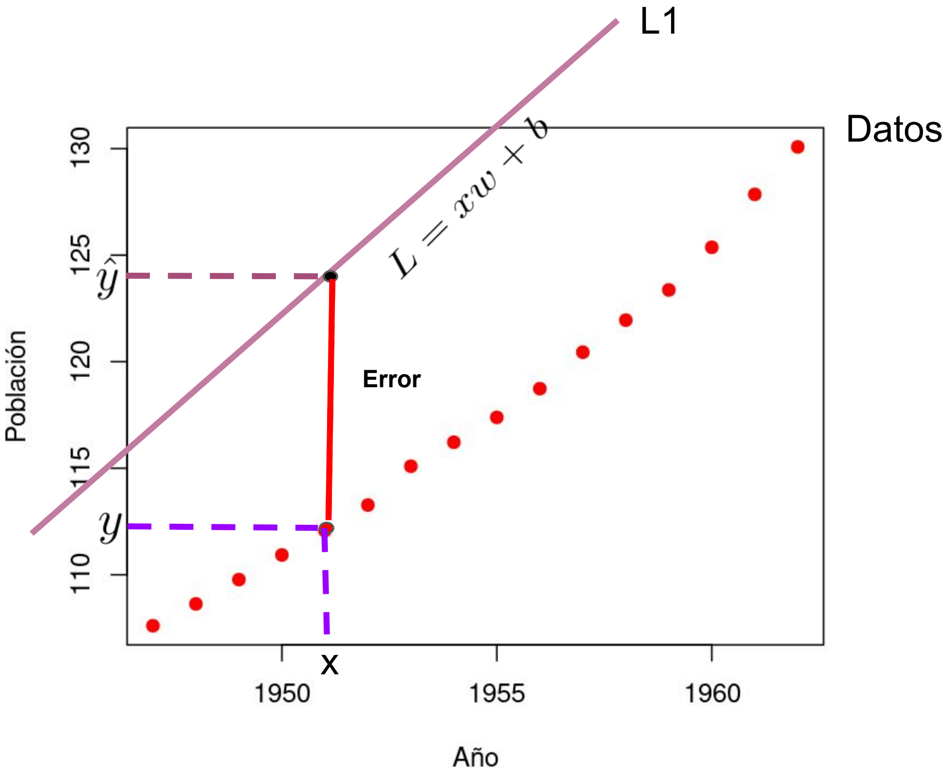


UTEC UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

How do I find the correct values of w and b?



How do I find the correct values of w and b?



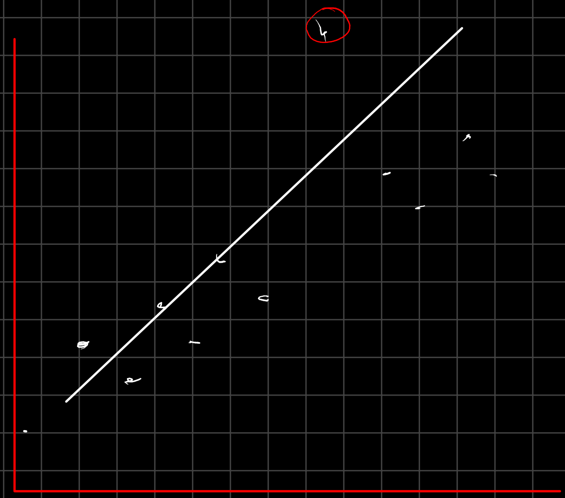
How do I find the correct values of w and b?

Hipótesis

$$h(y_i) = x_i w + b$$

Loss Function

$$\mathcal{L} = \frac{1}{2n}\sum_{i=0}^{n}(y_i - \hat{y_i})^2$$



MSE

$(y_i - \hat{y}_i)$

2 3 2 3 2 3 20

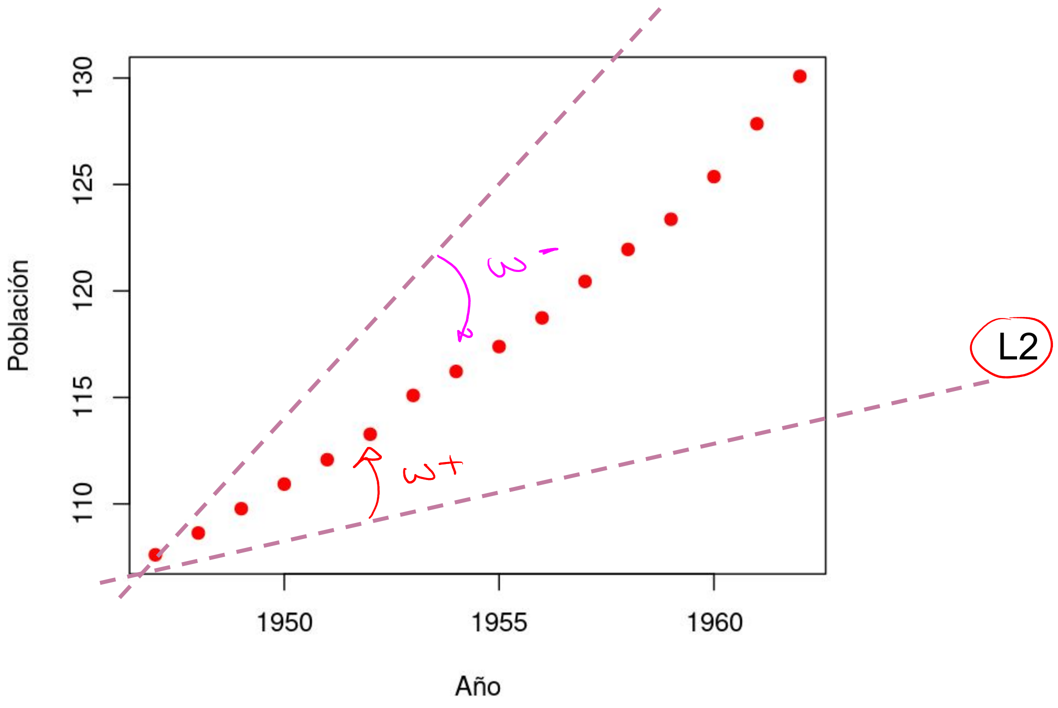
$\overline{}$

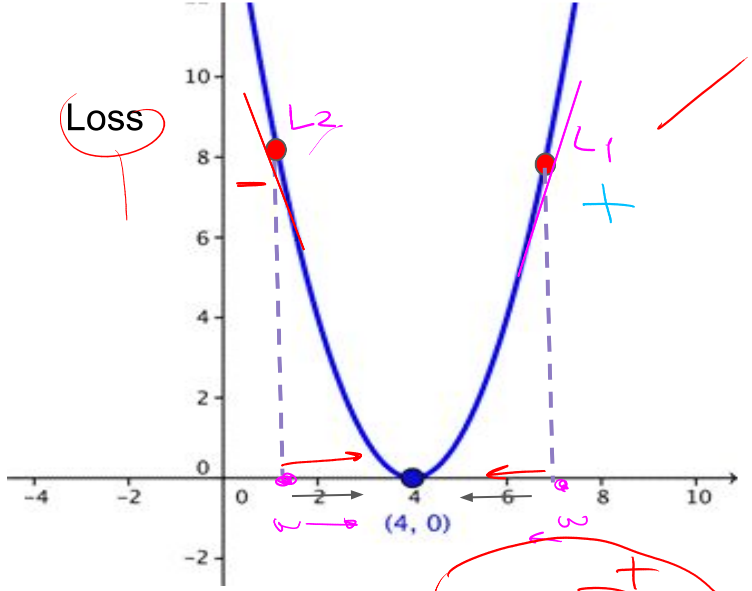
2.5

How do I find the correct values of w and b?

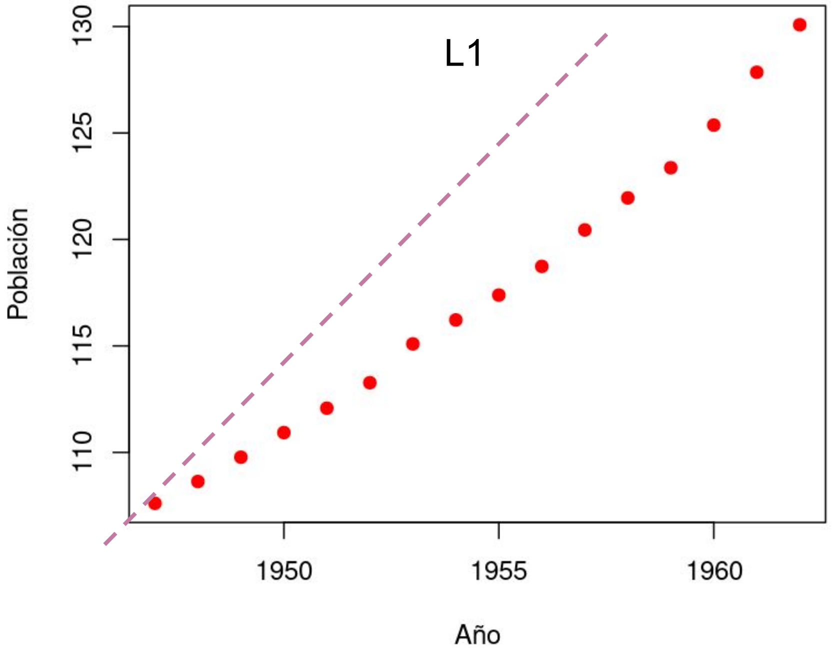
$$h(y_i) = x_i w$$

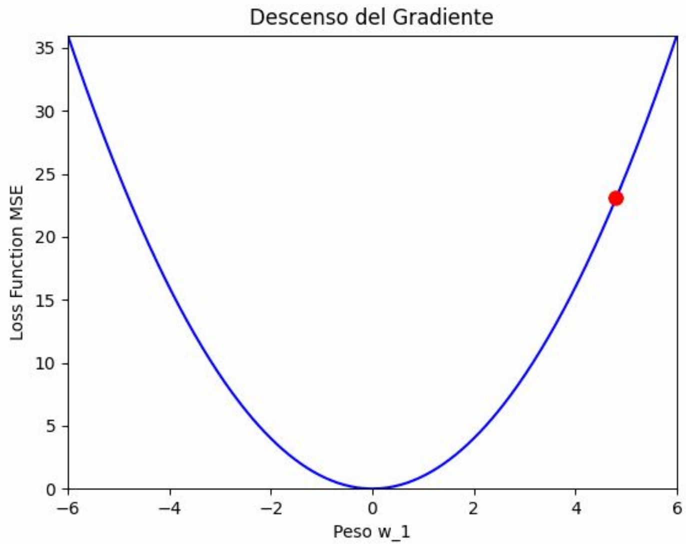
$W = w - \alpha \frac{\partial L}{\partial w}$

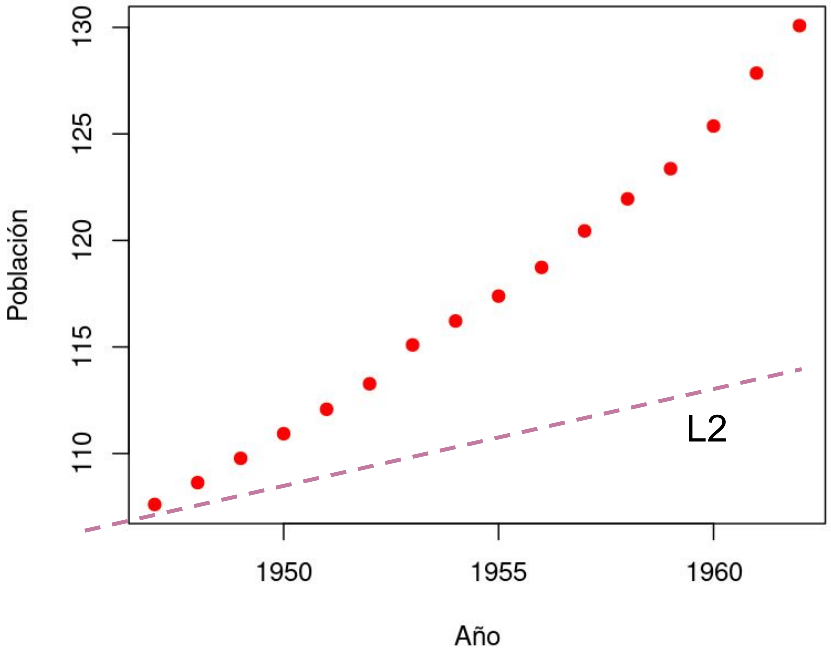


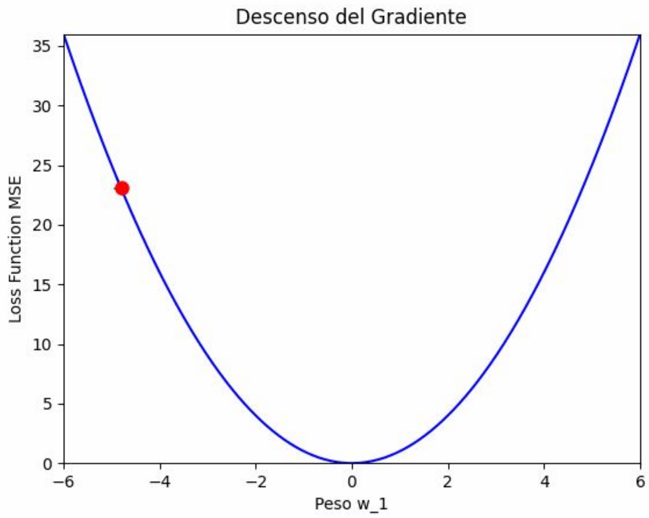


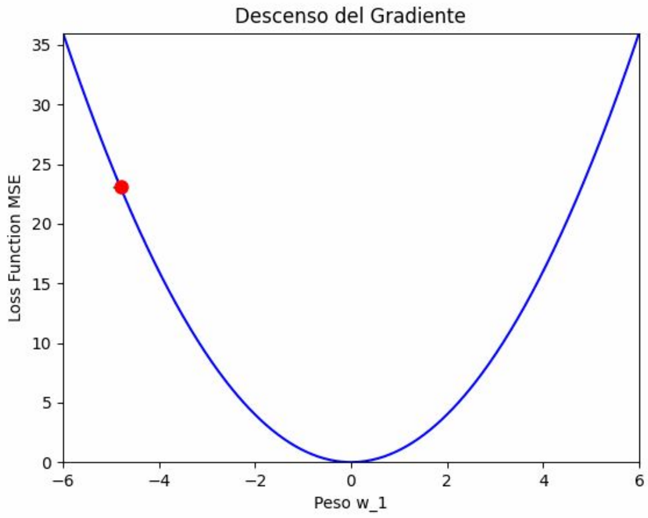
$w = w - \alpha \frac{\partial L}{\partial w}$


L1
130
125
120
115
110
Población
1950
1955
1960
Año

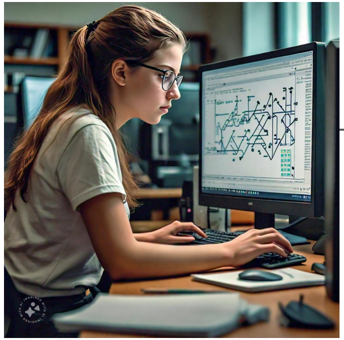

Descenso del Gradiente
35
30
25
20
15
10
5
0
Loss Function MSE
−6
−4
−2
0
2
4
6
Peso w_1


130
125
120
115
110
Población
1950
1955
1960
Año
L2


Descenso del Gradiente
35
30
25
20
15
10
5
0
Loss Function MSE
−6
−4
−2
0
2
4
6
Peso w_1



$$w = w - \alpha \frac{\partial Loss}{\partial w}$$

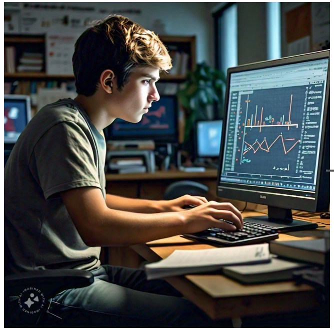


Training time

$$model = x_i w + b$$

$$loss = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{2n}$$

$$w = w - \alpha \frac{\partial Loss}{\partial w}$$

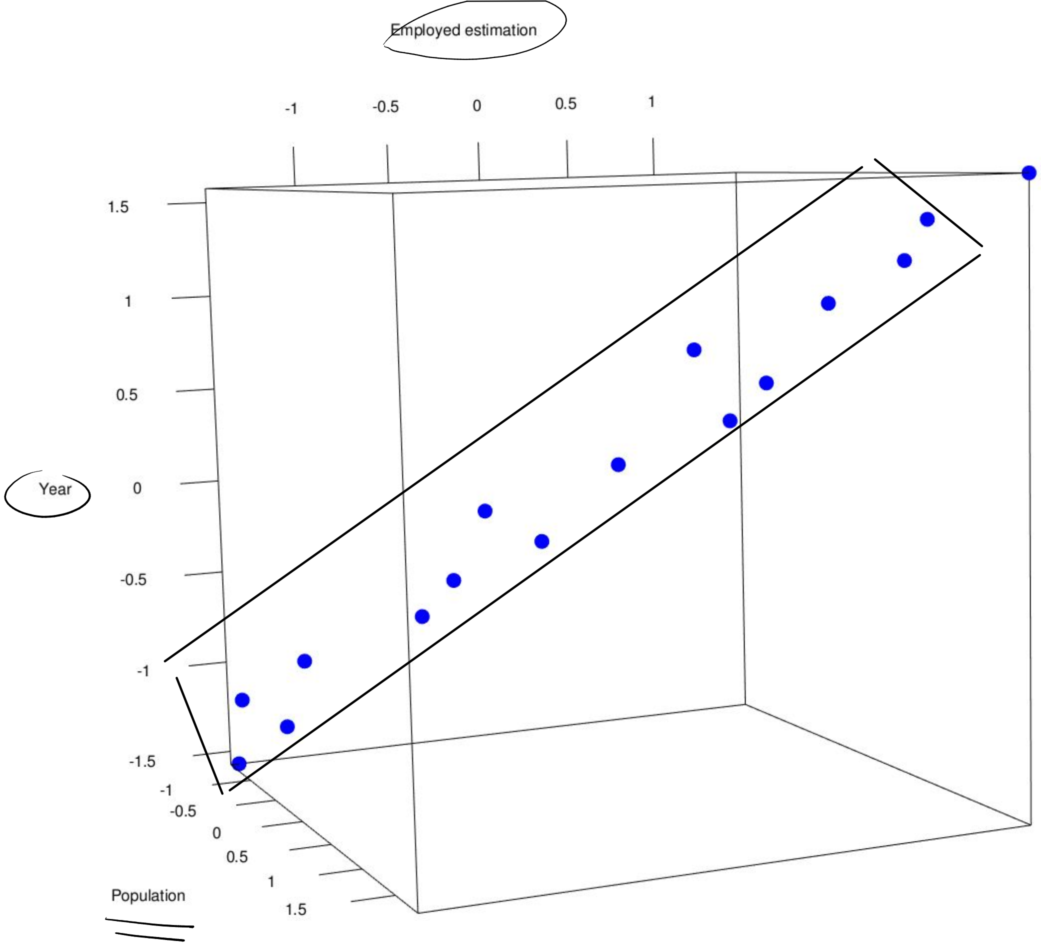


Inference time

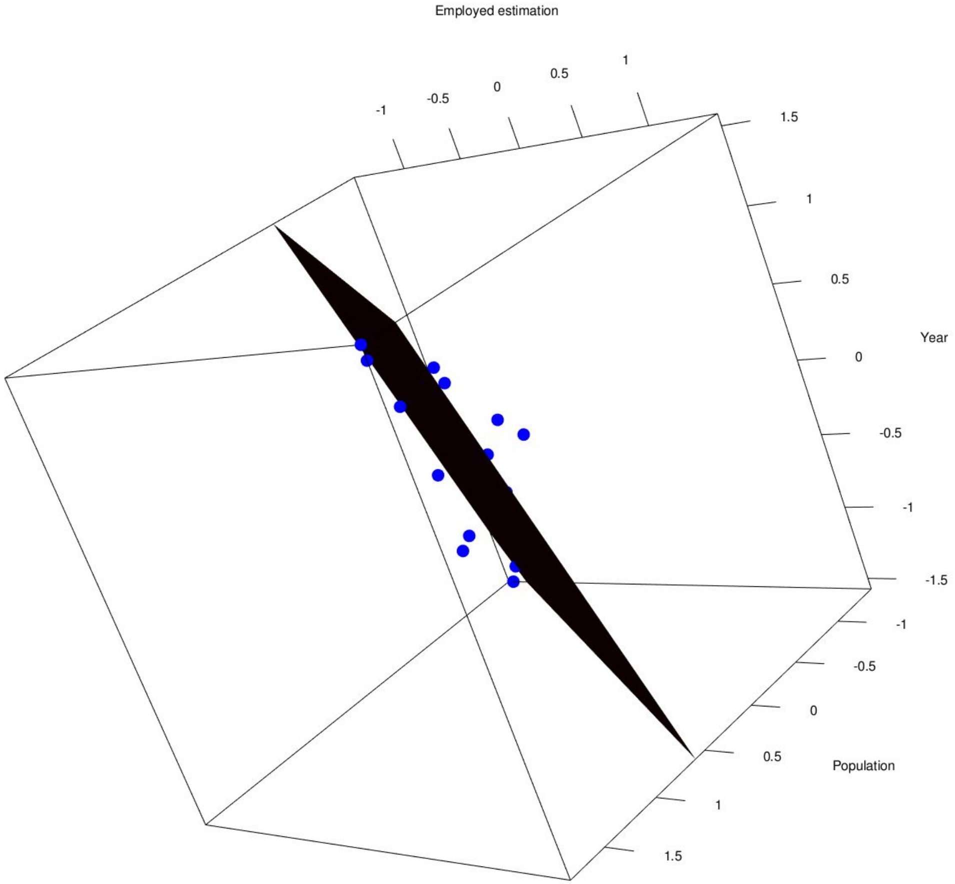
$$model = x_i w^* + b^*$$

Objective: Understand the idea behind multivariable linear regression

How do we approach this set of points to **predict employability** based on **population** and **year**?



How can we **create** the **plane** that **best fits** this **set of points**?



UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

Plane Equation

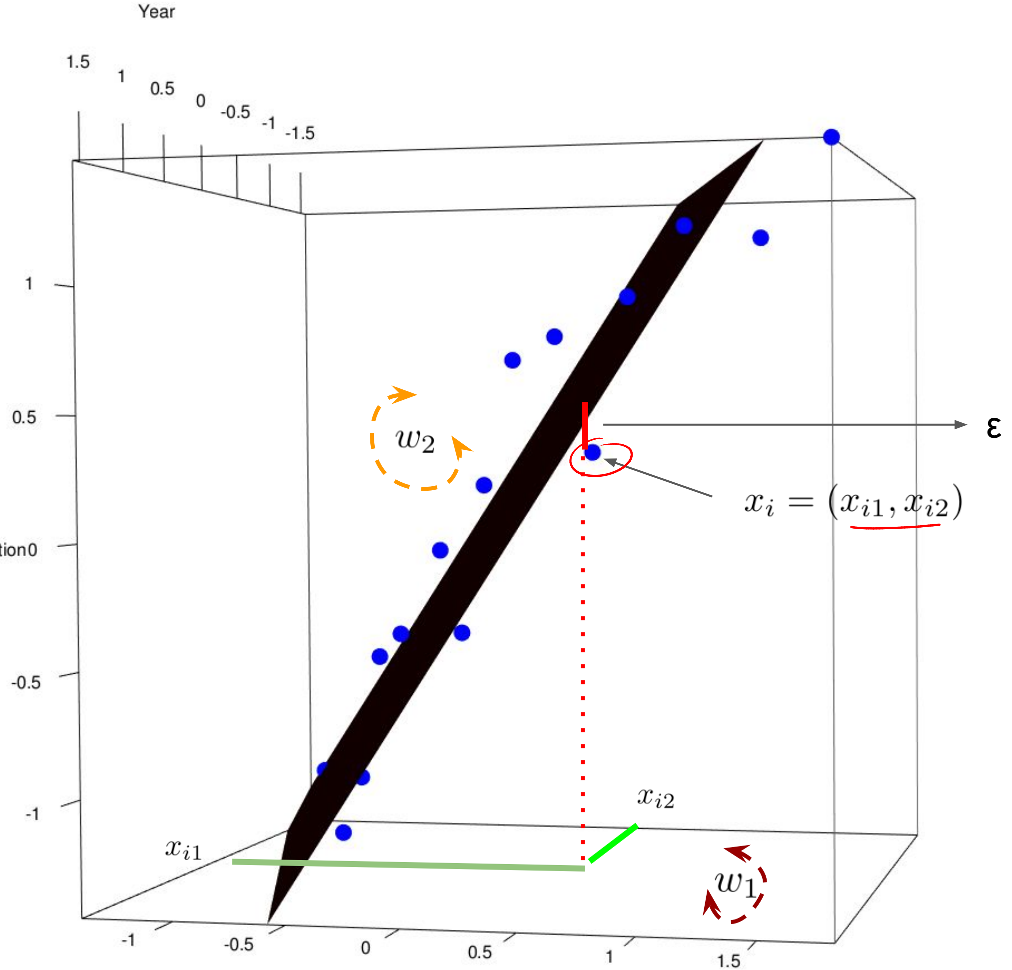
angulo angulo desplaz

$$ax_1 + bx_2 + c = 0 \longrightarrow x_1 w_1 + x_2 w_2 + b = 0$$

Area | # Pisos | costo

x_i 200 2 300$

$\boxed{x_{i1} \mid x_{i2}}$ Y_i



recta

Hypothesis for **univariate** linear regression

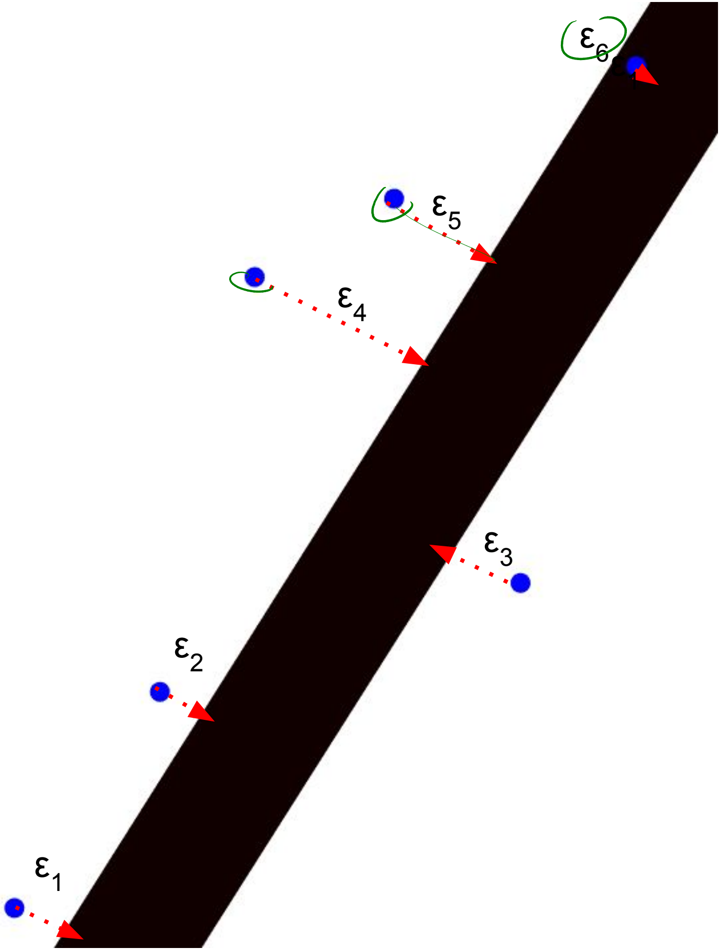
$$h(x_i) = x_i w + b$$

plano

Hypothesis for **multivariate** linear regression

$$h(x_i) = x_{i1} w_1 + \boxed{x_{i2} w_2} + b$$

$$h(x_i) = x_{i1}w_1 + x_{i2}w_2 + b$$



$$\mathcal{L} = \frac{\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_6}{6}$$

$$\mathcal{L} = \frac{\sum_{i=1}^{n} \varepsilon_i}{n}$$

MSE

MAE

What would be missing?

Hypothesis $$h(x_i) = x_{i1}w_1 + x_{i2}w_2 + b$$

Loss Function $$\mathcal{L} = \frac{\sum_{i=1}^{n}(y_i - h(x_i))^2}{2n}$$

Derivatives $$\frac{\partial Loss}{\partial b} \quad \frac{\partial Loss}{\partial w_1} \quad \frac{\partial Loss}{\partial w_2}$$

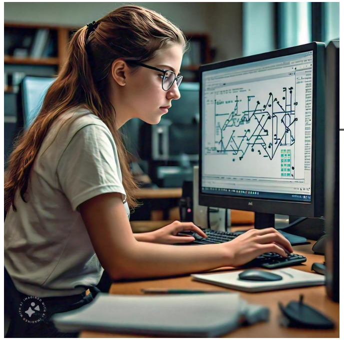
Change parameters $$w_i = w_i - \alpha\frac{\partial loss}{\partial w_i}$$

$$L = \sum_{i=0}^{n} \frac{(y_i - \boxed{h(x_i)})^2}{2n}$$

$$\frac{\partial L}{\partial b} = \sum_{i=0}^{n} \frac{(y_i - h(x_i))}{n} \cdot (-1)$$

$$\frac{\partial L}{\partial w_1} = \sum_{i=0}^{n} \frac{(y_i - h(x_i))}{n} (-x_{i1})$$

$$\frac{\partial L}{\partial w_2} = \sum_{i=0}^{n} \frac{(y_i - h(x_i))}{n} (-x_{i2})$$

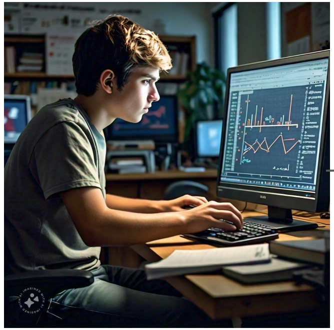


Training time

$$model = x_{i1}w_1 + \boxed{x_{i2}w_2} + b$$

$$loss = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{2n}$$

$$w = w - \alpha\frac{\partial Loss}{\partial w}$$



Inference time

$$model = x_{i1}w_1^* + x_{i1}w_2 + b^*$$

(b, w_1, w_2) — pesos

$$\underbrace{\begin{array}{c|ccc} & F_0 & F_1 & F_2 \\ \hline x_1 & 1 & x_{11} & x_{12} \\ x_2 & 1 & & \\ x_3 & 1 & & \\ \vdots & & & \\ x_i & 1 & x_{i1} & x_{i2} \\ & & & \\ x_n & 1 & & \end{array}}_{X\_train} \ * \ \underbrace{\begin{array}{c|c} b & \\ w_1 & \\ w_2 & \end{array}}_{w^t} \ = \ \underbrace{\begin{bmatrix} \\ \\ \vdots \\ \\ \\ \end{bmatrix}}_{\hat{y}}$$

$W_{n \times 2}$ $\qquad$ $W_{3 \times 1}$ $\qquad$ $n \times 1$

* Armadillo
* eisen

$$h(x_i) = x_{i1} w_1 + x_{i2} w_2 + b$$

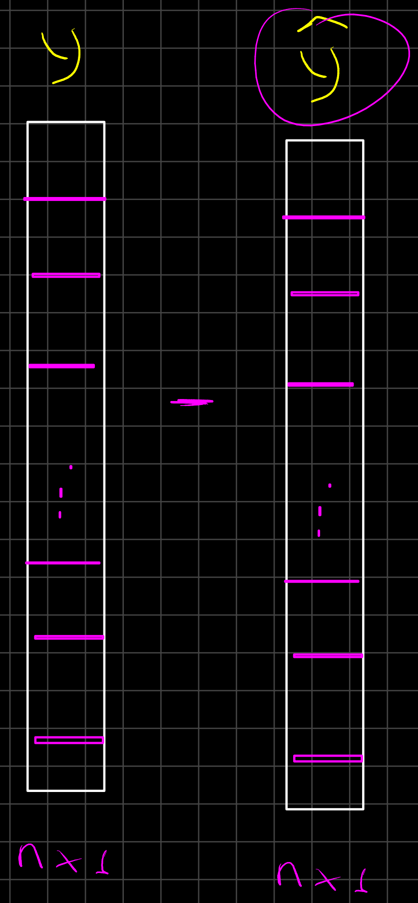
$$\begin{array}{|c|c|c|c|} \hline 1 & b & w_1 & w_2 \\ \hline \end{array}$$

$$L = \sum_{i=0}^{n} (y_i - h(x_i))^2 / 2n$$

$$\frac{\partial L}{\partial b} = \sum_{i=0}^{n} \frac{(y_i - h(x_i))}{n} (-1)$$

$$\frac{\partial L}{\partial w_1} = \sum_{i=0}^{n} \frac{(y_i - h(x_i))}{n} (-x_{i1})$$

$$\frac{\partial L}{\partial w_2} = \sum_{i=0}^{n} \frac{(y_i - h(x_i))}{n} (-x_{i2})$$

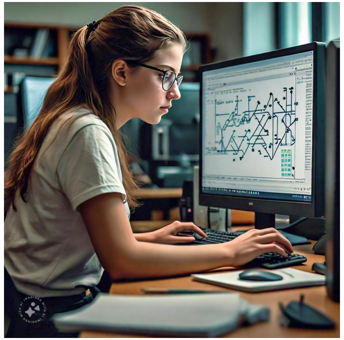


$$\text{sum}(Y - Y_{\text{pred}}) / n$$

Video

Generalizing to a k-dimensional space

UTEC UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA



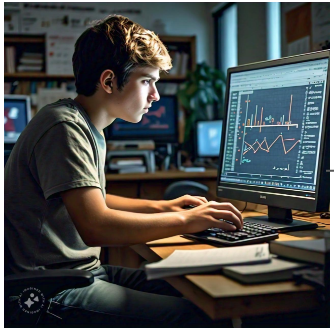
Training time

$$model = x_{i1}w_1 + x_{i1}w_2 + \ldots + x_{ik}w_k + b$$

$$loss = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{2n}$$

$$w = w - \alpha\frac{\partial Loss}{\partial w}$$

$$\frac{\partial Loss}{\partial w_i} = \frac{1}{n}\sum_{i=0}^{n}(y_i - h(x_i))^2$$



Inference time

$$model = x_{i1}w_1 + x_{i1}w_2 + \ldots + x_{ik}w_k + b$$

```
def train(x, y, umbral, alfa):
    w = [np.random.rand() for i in range(1:k)]
    b = np.random.rand()
    L = Error(x, y, w, b)
    loss = []
    while (L > umbral):
        db, dw = derivada(x, y, w, b)
        b, w = update(w, b, alfa, db, dw)
        L = Error(x, y, w, b)
        print(L)
        loss.append(L)
    return b, w
```

$$h(x_i) = x_{i1} w_1 + x_{i2} w_2 + \cdots + x_{ik} w_k + b$$

F_1 Area | F_2 N. pisos | F_3 ... N cuartos | F_k N baños ...

x_1

x_2

$\vdots$

$x_i = \boxed{x_{i1} \quad x_{i2} \quad x_{i3} \quad \cdots \quad x_{nf}}$

x_n

$$h(X_i) = X_{i1} w_1 + X_{i2} w_2 + \cdots + X_{ik} w_k + b$$

$$h(X_i) = b + X_{i1} w_1 + X_{i2} w_2 + \cdots + X_{ik} w_k$$

$$h(X_i) = w_0 1 + X_{i1} w_1 + X_{i2} w_2 + \cdots + X_{ik} w_k$$

$$h(X_i) = [1 \quad X_{i1} \quad X_{i2} \cdots X_{ik}] \cdot [w_0 \quad w_1 \cdots w_k]^t$$

$$\boxed{h(x_i) = X_i w^t}$$

$\rightarrow$ modelo de Regresión Lineal

$$\boxed{h(x_i) = x_i w^t} \leftarrow \text{Transpuesta}$$

$$L = \sum_{i=1}^{n} \frac{(y_i - h(x_i))^2}{2n} \text{ — MSE}$$

$$\frac{\partial L}{\partial w_0} = \sum_{i=1}^{n} \frac{(y_i - h(x_i))}{n} (-1)$$

$$\frac{\partial b}{\partial w_1} = \sum_{i=1}^{n} \frac{(y_i - h(x_i))}{n} (-x_{i1})$$

$$\frac{\partial b}{\partial w_2} = \sum_{i=1}^{n} \frac{(y_i - h(x_i))}{n} (-x_{i2})$$

$$\vdots$$

$$\frac{\partial b}{\partial w_j} = \sum_{i=1}^{n} \frac{(y_i - h(x_i))}{n} (-x_{ij})$$

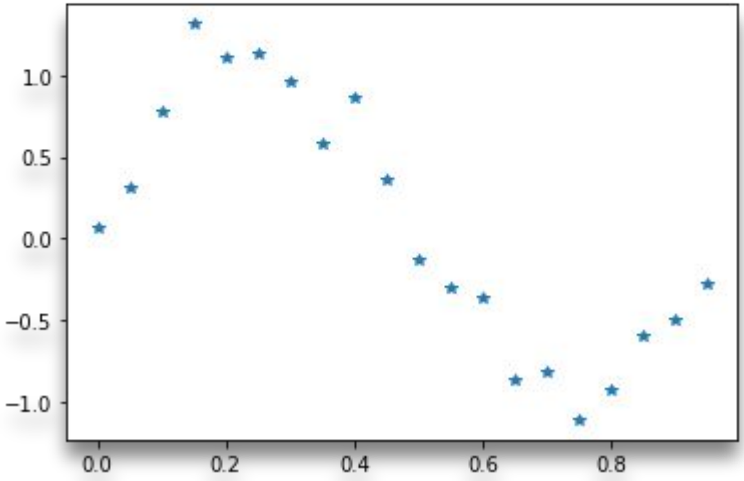
$$\frac{\partial L}{\partial w_j} = \sum_{i=1}^{n} \frac{(y_i - h(x_i))}{n} \frac{\partial}{\partial w_j} \left[\cancel{y_i} - \left(\cancel{w_0} + \cancel{x_{i1} w_1} + \cdots + x_{ij} w_j + \cdots \cancel{x_{in} w_n} \right) \right]$$

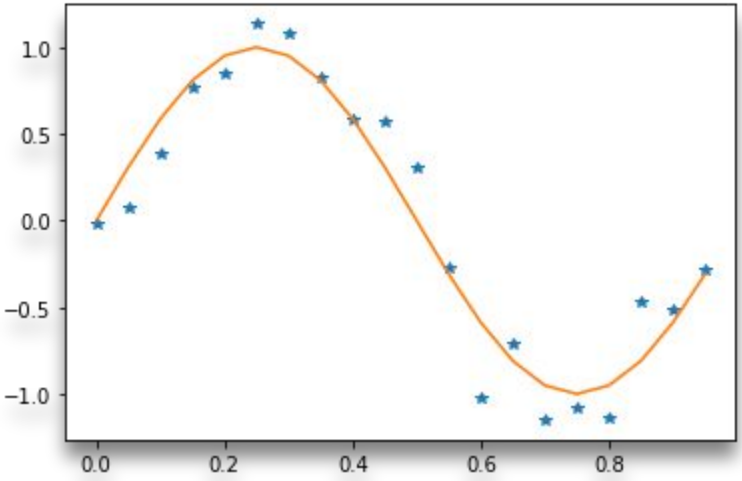
$$\frac{\partial L}{\partial w_j} = \sum_{i=1}^{n} \frac{(y_i - h(x_i))}{n} \left(- x_{ij} \right)$$

$$\boxed{w_j = w_j - \alpha \frac{\partial L}{\partial w_j}}$$

Linear Regression

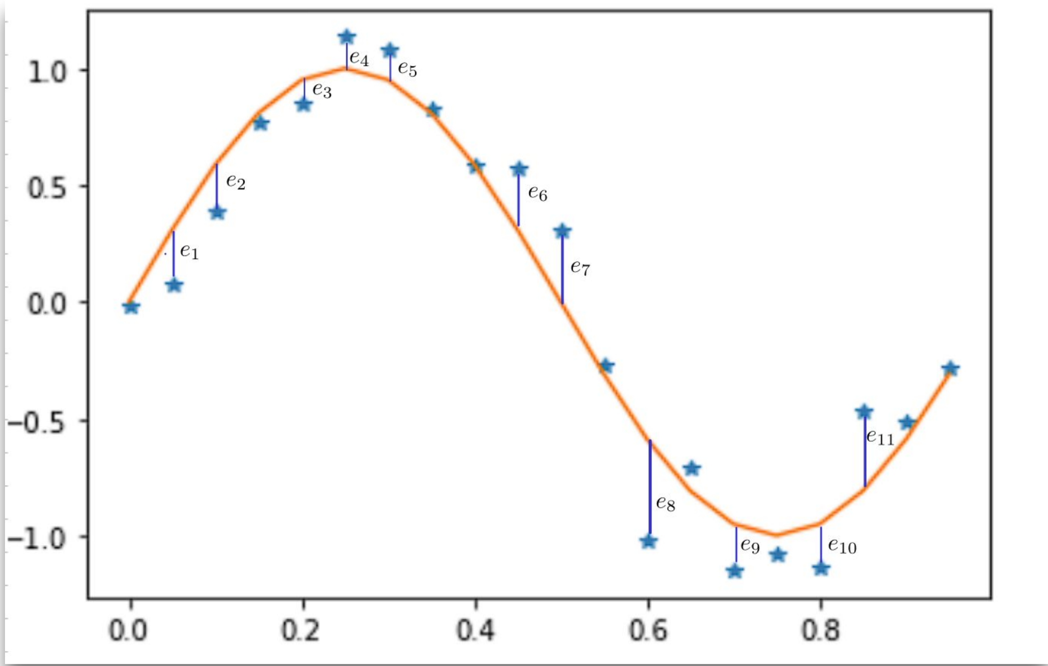
Objective: Understand the idea behind nonlinear


1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8


1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8

Loss Function

Loss



Epochs

Loss Function

$$\mathcal{L} = \frac{\sum_{i=0}^{n}(y_i - h(x_i))^2}{2n}$$

UTEC UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

What would be missing?

Hiper parámetro

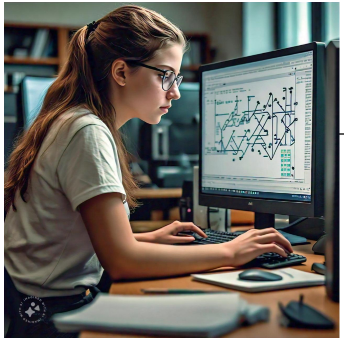
grado del polinomio

Hypothesis : $$h(x_i) = b + x_i w_1 + x_i^2 w_2 + x_i^3 w_2 + \ldots + + x_i^p w_2$$

Loss Function : $$\mathcal{L} = \frac{\sum_{i=1}^{n} (y_i - h(x_i))^2}{2n}$$

MSE

Change parameters → Find derivatives of the error with respect to the parameters



Training time

$$modelo = x_i w_1 + x_i^2 w_2 + \dots + x_i^k w_k + b$$

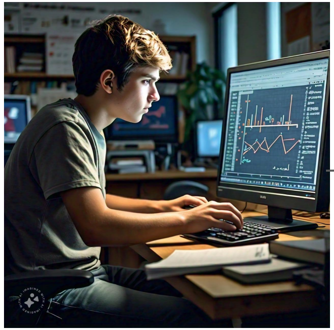
Real

Pred

$$loss = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{2n}$$

$$w = w - \alpha \frac{\partial Loss}{\partial w}$$

$$\frac{\partial Loss}{\partial w_i} = \frac{1}{n} \sum_{i=0}^{n} (y_i - h(x_i))^2$$



Inference time

$$model = x_{i1} w_1 + x_{i1} w_2 + \dots + x_{ik} w_k + b$$

$$h(x_i) = b + x_i w_1 + x_i^2 w_2 + \dots + x_i^p w_p$$

$$h(x_i) = x_i^0 w_0 + x_i w_1 + x_i^2 w_2 + \dots + x_i^p w_p$$

donde $b = w_0$

$$= [x_i^0 \; x_i^1 \; x_i^2 \dots x_i^p][w_0 \; w_1 \dots w_p]^t$$

$$\boxed{h(x_i) = X_i W^t}$$ ← hipótesis

MSE $\|Y - XW^t\|_2^2$

$$L = \frac{\sum_{i=0}^{n} (y_i - h(x_i))^2}{2n} = \|Y - XW^t\|_2^2$$

$$L = \|Y - XW^t\|_2^2$$ ← Norma L_2

$\|Y\|_2 \rightsquigarrow$ norme L_2 $\quad \left(\cancel{\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}}\right)^2$

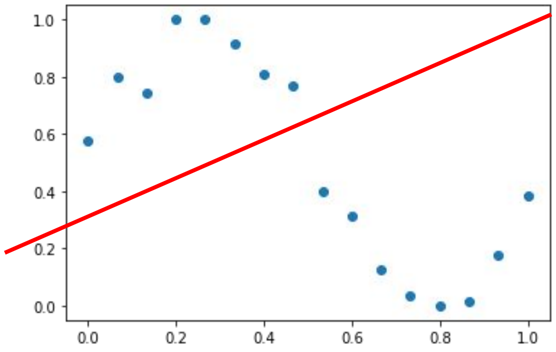
$$\frac{\partial L}{\partial w_0} = \sum_{i=0}^{n} \frac{\left(y_i - h(x_i)\right)}{n} \frac{\partial}{\partial w_0}\left[y_i - \left[w_0 + x_i w_1 + x_i^2 w_2 + \cdots x_i^p w_p \right] \right]$$

$$\frac{\partial L}{\partial w_0} = \sum_{i=0}^{n} \frac{\left(y_i - h(x_i)\right)}{n} (-1)$$

$$\frac{\partial L}{\partial w_j} = \sum_{i=0}^{n} \frac{\left(y_i - h(x_i)\right)}{n} \frac{\partial}{\partial w_j}\left(\cancel{y_i} - \left[\cancel{w_0} + \cancel{x_i w_1} + \cancel{x_i^2 w_2} + \cdots + x_i^j w_j + \cdots \right] \right)$$

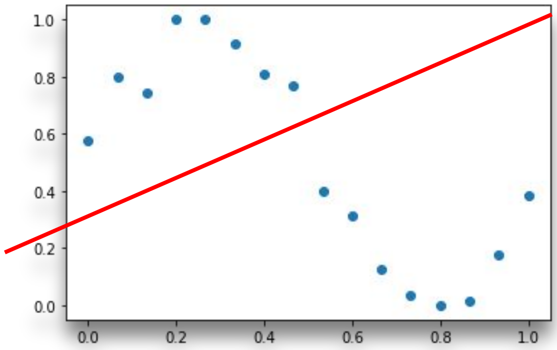
$$\frac{\partial L}{\partial w_j} = \sum_{i=0}^{n} \frac{\left(y_i - h(x_i)\right)}{n} \left(- x_i^j\right)$$

Underfitting



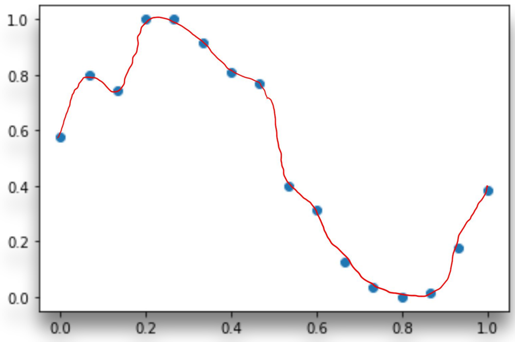
$$h(x_i) = x_i w + b$$

Underfitting



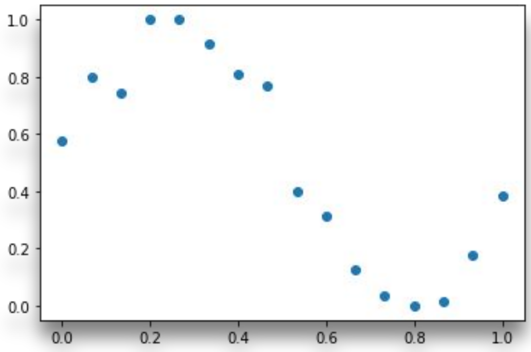
$$h(x_i) = x_i w + b$$

Overfitting



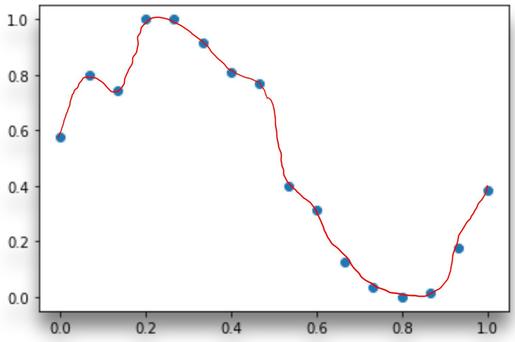
$$h(x_i) = x_i^0 w_0 + x_1^1 w_1 + \ldots + x_i^{20} w_{20}$$

Underfitting



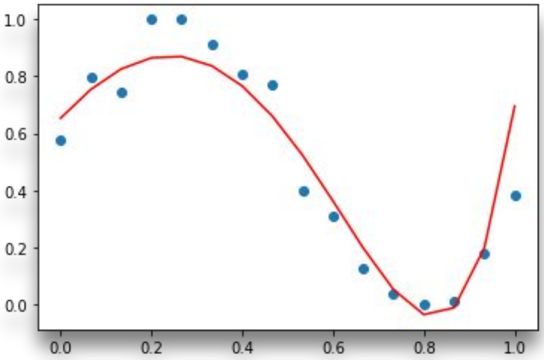
$$h(x_i) = x_i w + b$$

Overfitting



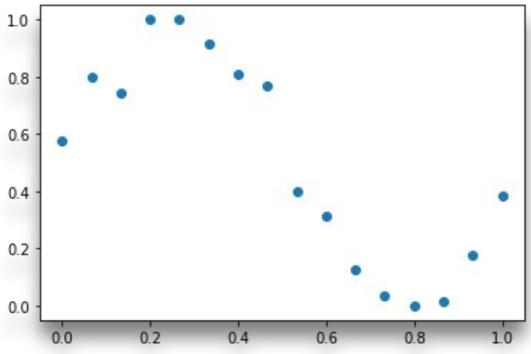
$$h(x_i) = x_i^0 w_0 + x_1^1 w_1 + \ldots + x_i^{20} w_{20}$$

good



$$h(x_i) = x_i^0 w_0 + x_1^1 w_1 + \ldots + x_i^3 w_3$$

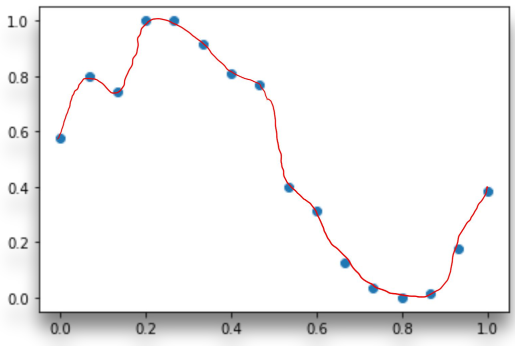
Underfitting



$$h(x_i) = x_i w + b$$

- Simple Model
- Low Capacity Model

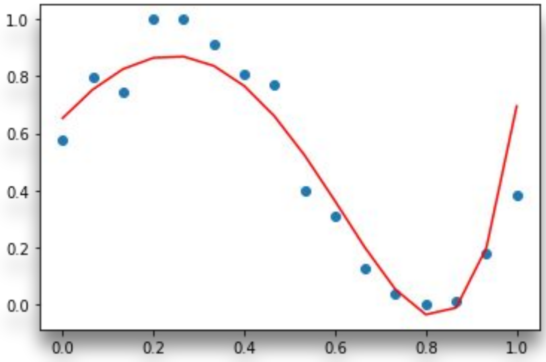
Overfitting



$$h(x_i) = x_i^0 w_0 + x_1^1 w_1 + \ldots + x_i^{20} w_{20}$$

- Highly Complex Model
- High Capacity Model

good



$$h(x_i) = x_i^0 w_0 + x_1^1 w_1 + \ldots + x_i^3 w_3$$

- Data-Fitting Model
- Model with Adequate Capacity

Linear No Regression