Regresión Lineal

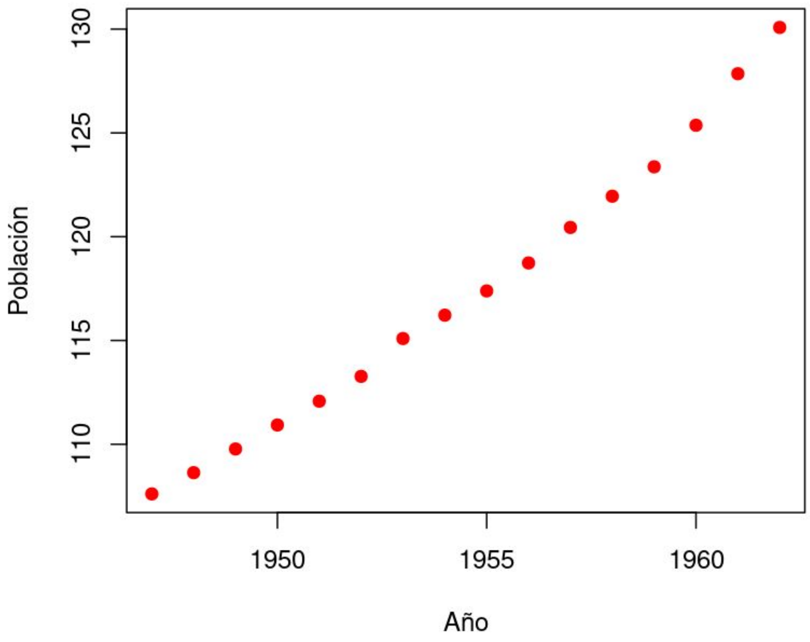
Cristian López Del Alamo
clopezd@utec.edu.pe
IPRODAM3D - Research group

2022

1 Regresión Lineal univariada

Objetivo: Entender la matemática detrás de la regresión lineal univariada

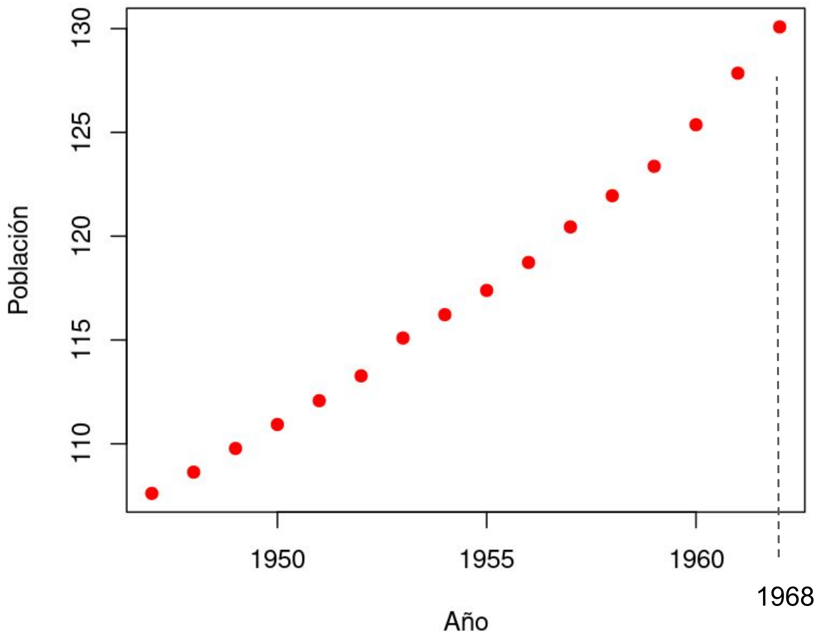
Predicción de la población

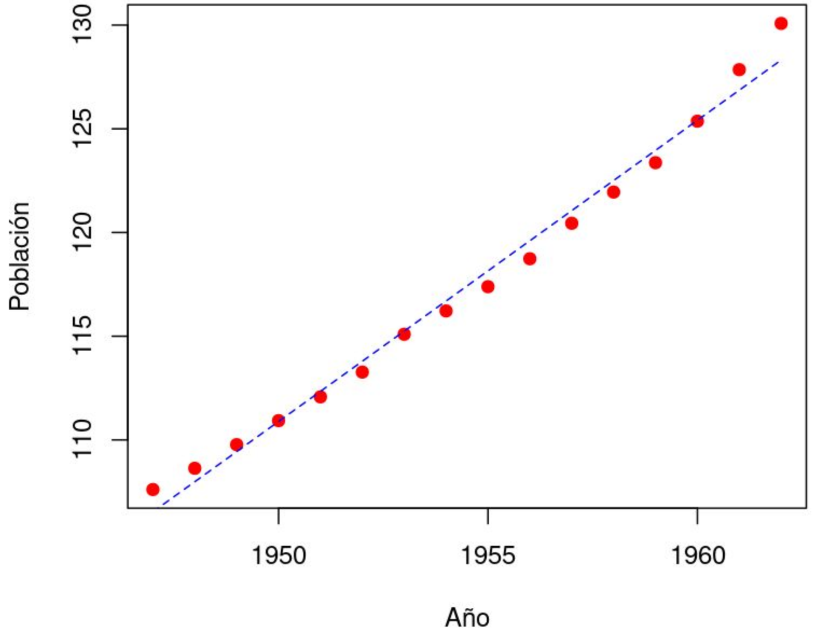


	Population	Year
1947	107.608	1947
1948	108.632	1948
1949	109.773	1949
1950	110.929	1950
1951	112.075	1951
1952	113.270	1952

…

Predicción de la población





Modelo

$$L = xw + b$$

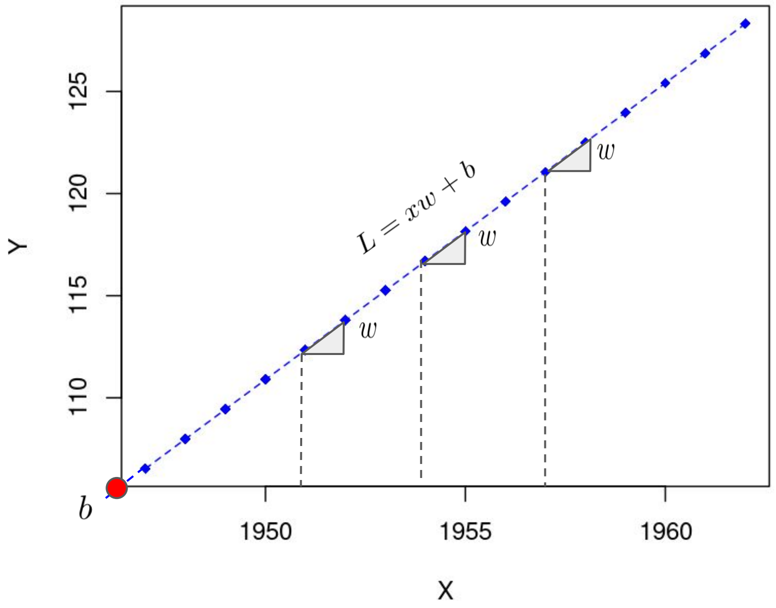
Parámetros

Modelo

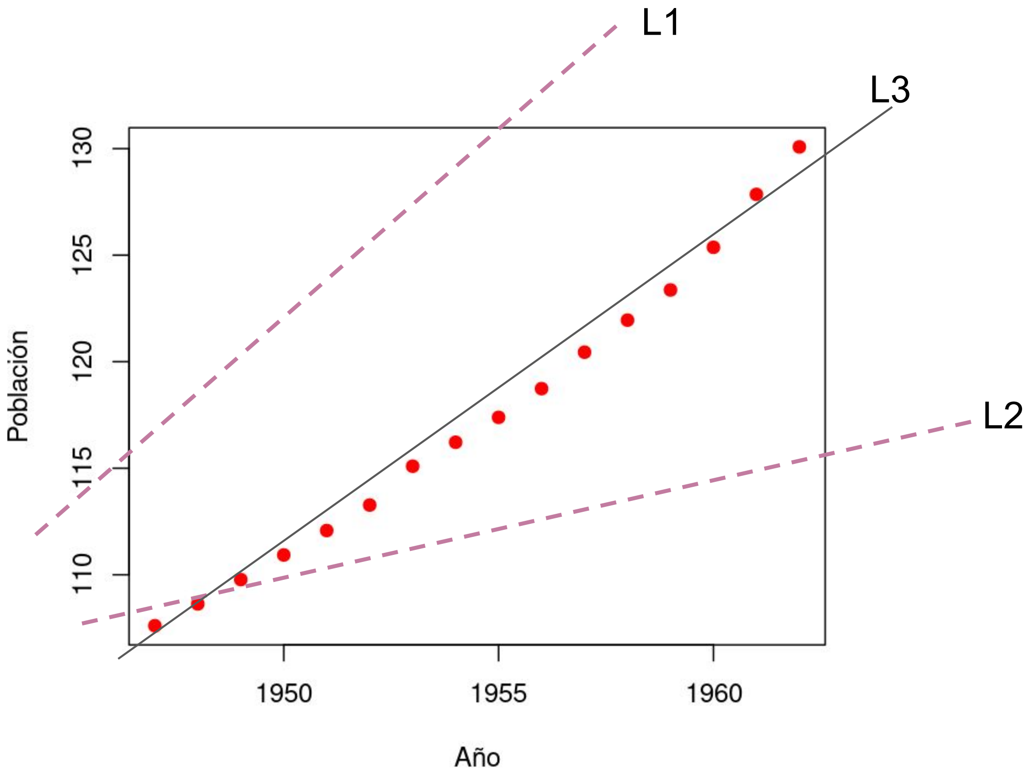
$$L = xw + b$$

bias

Pendiente

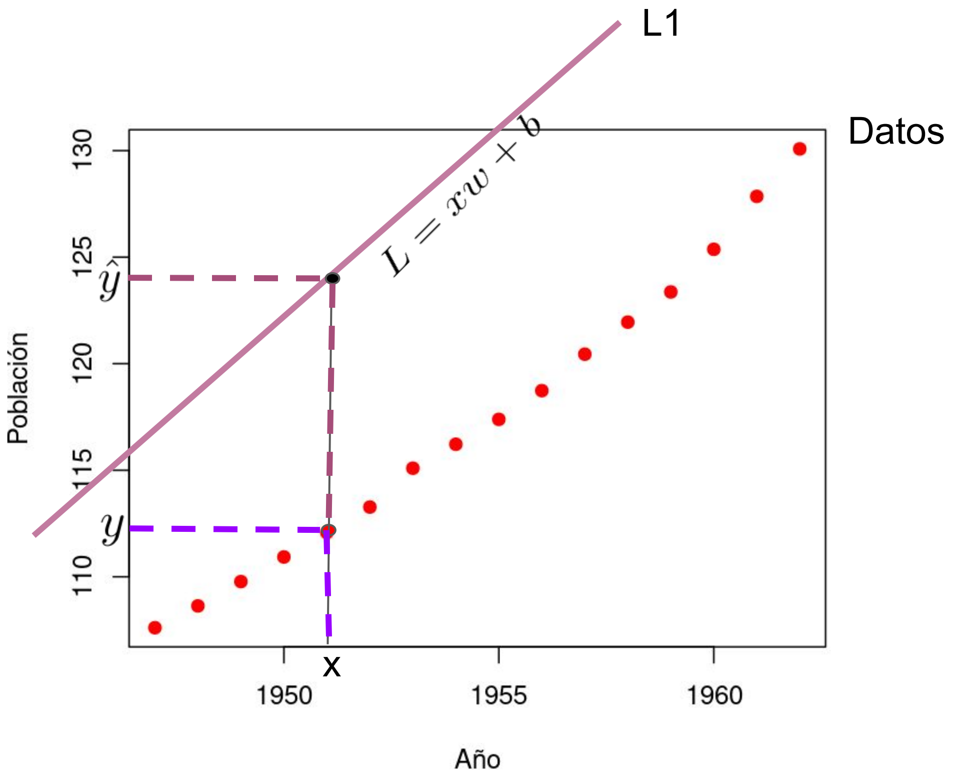


¿Cómo encuentro los valores correctos de w y b?



$$L = xw + b$$

¿Cómo encuentro los valores correctos de w y b?



$$e_i = (y_i - \hat{y}_i)^2$$

$$\hat{y}_i = x_i w + b$$

$$obj : min(\frac{1}{n}\sum_{i=0}^{n} e_i)$$

$$obj : \min \frac{1}{2n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$

¿Cómo encuentro los valores correctos de w y b?

Hipótesis

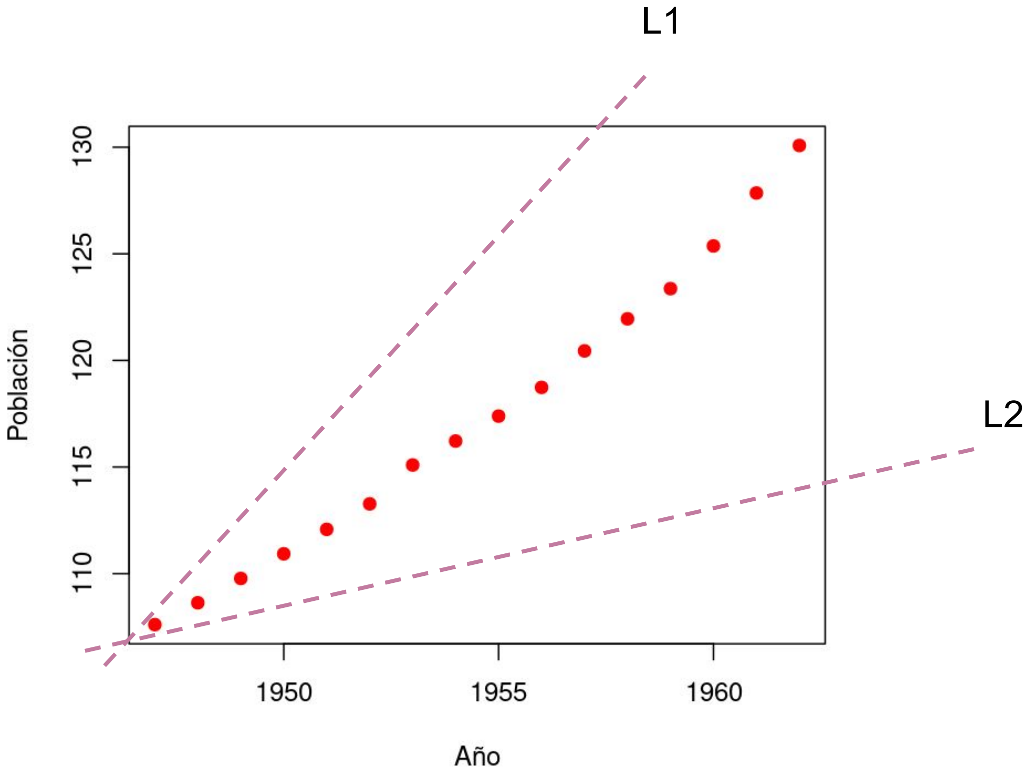
$$h(y_i) = x_i w + b$$

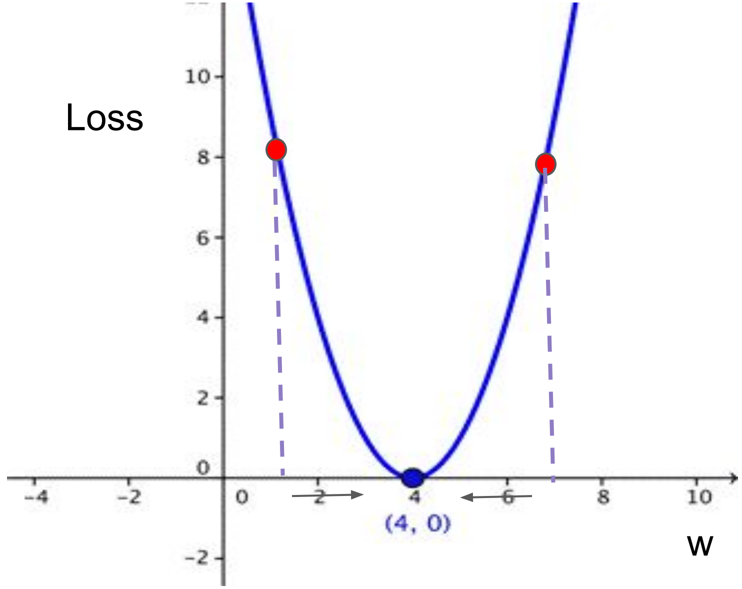
Loss Function

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

$$\mathcal{L} = \frac{1}{2n}\sum_{i=0}^{n}(y_i - (x_i w + b))^2$$

¿Cómo encuentro los valores correctos de w y b?





Cálculo de W y b

Cálculo de W y b

Cálculo de W y b

En esta sección se presentó la regresión lineal univariado, así como un modelo basado en aprendizaje de máquinas.

La siguiente clase veremos una implementación en pythoy para la regresión Lineal Univariada.

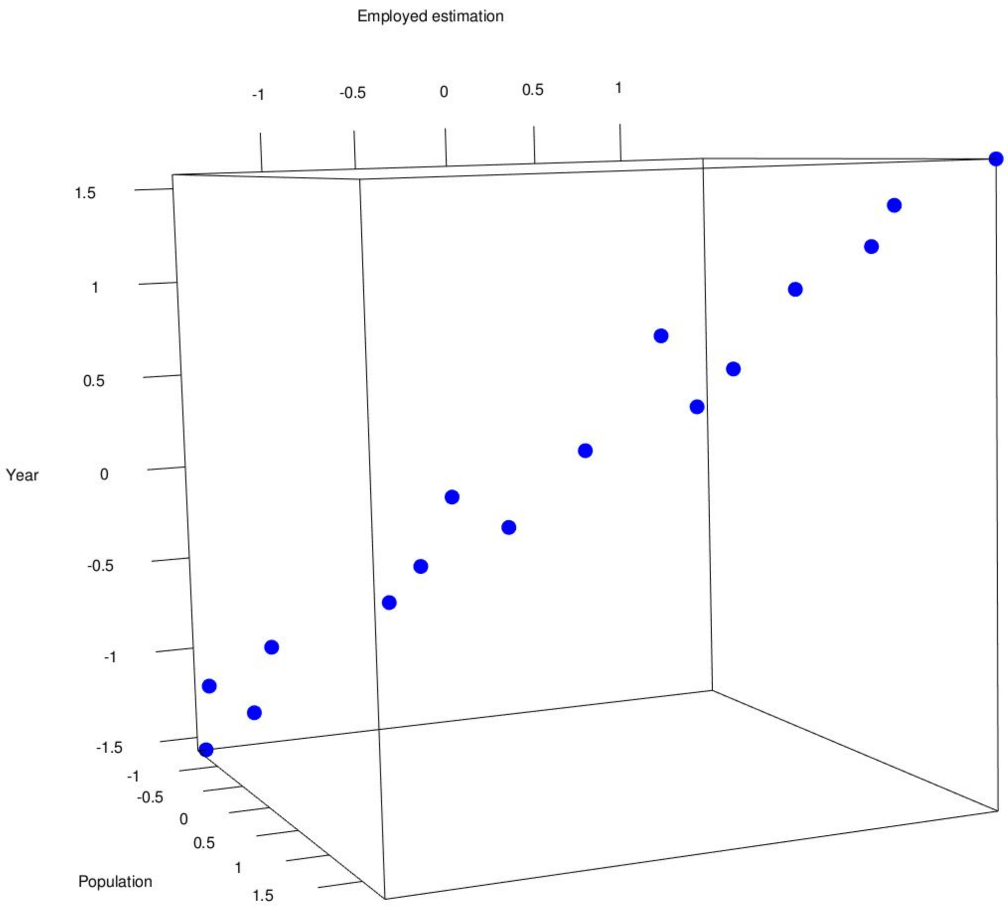
2 Regresión Lineal multivariable

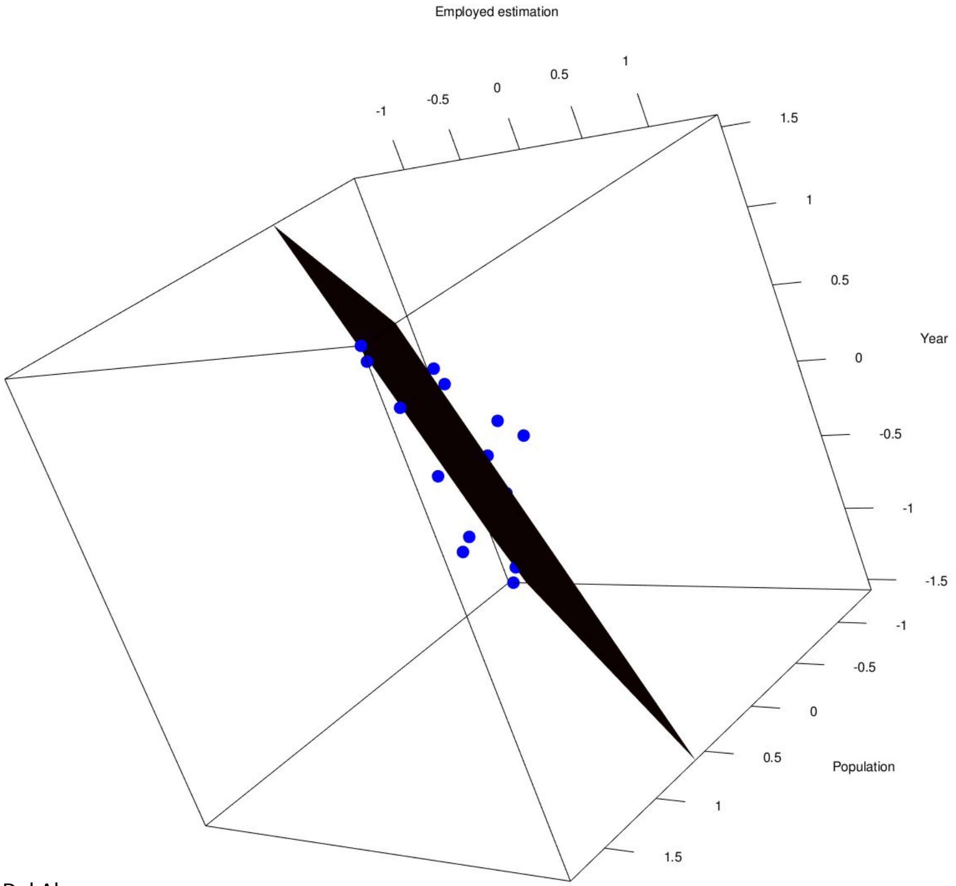
Objetivo: Entender la matemática detrás de la regresión lineal multivariable

```
data(longley)
print(names(longley))
plot(longley)

db<-longley[c("Year","Population","Employed")]
head(db)
db <- as.data.frame(scale(db))

fit <- lm(db$Employed~., db)
```

Year	Population	Employed
1947	107.608	60.323
1948	108.632	61.122
1949	109.773	60.171
1950	110.929	61.187
1951	112.075	63.221
1952	113.270	63.639

¿Cómo nos aproximamos a este conjunto de puntos para predecir la empleabilidad según población y año?.



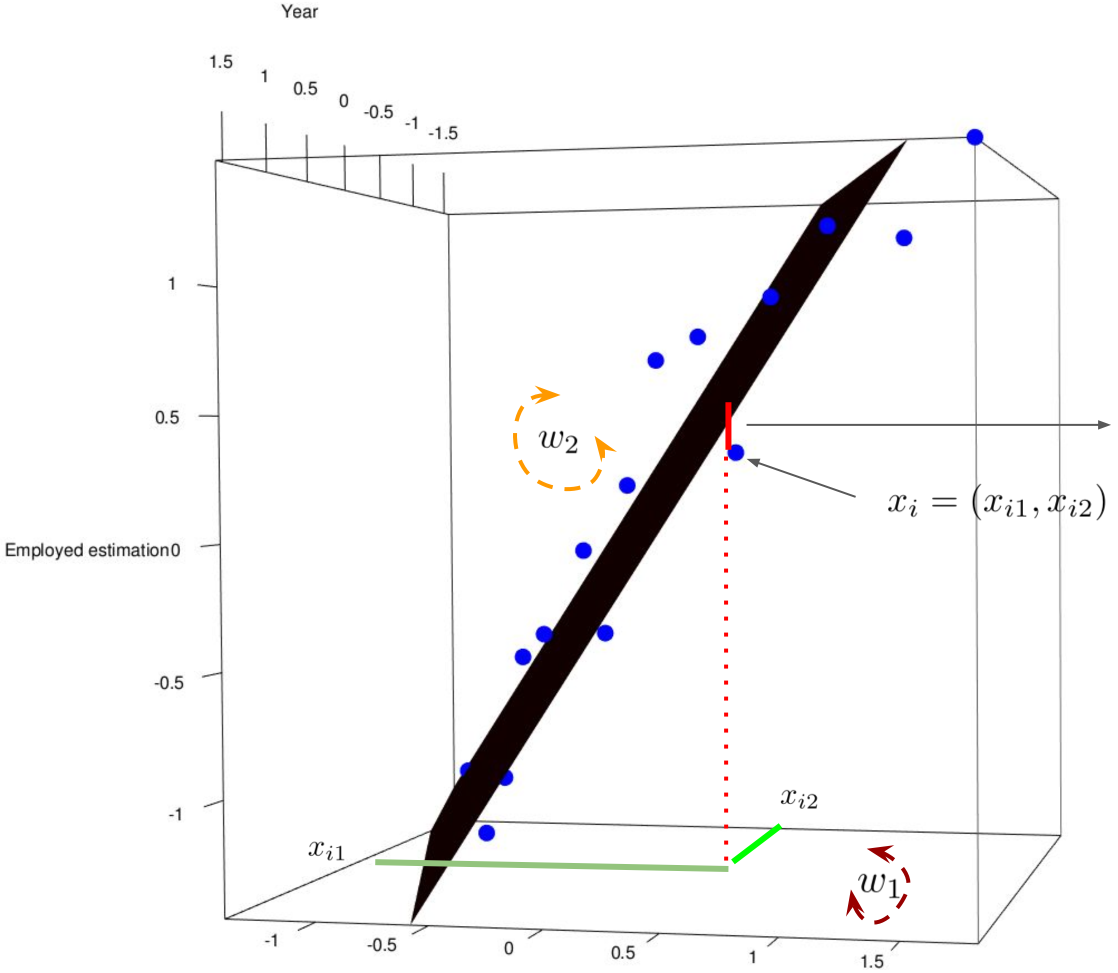


¿Cómo podemos crear el plano que mejor se ajuste a este conjunto de puntos?

Ecuación del plano

$$ax_1 + bx_2 + c = 0$$

$$x_1 w_1 + x_2 w_2 + b = 0$$



$\mathcal{L}(y_i, \hat{y}_i)$

$\mathcal{L}(y_i, h(x_i)$

$\mathcal{L}(y_i, x_{i1}w_1 + x_{i2}w_2 + b)$

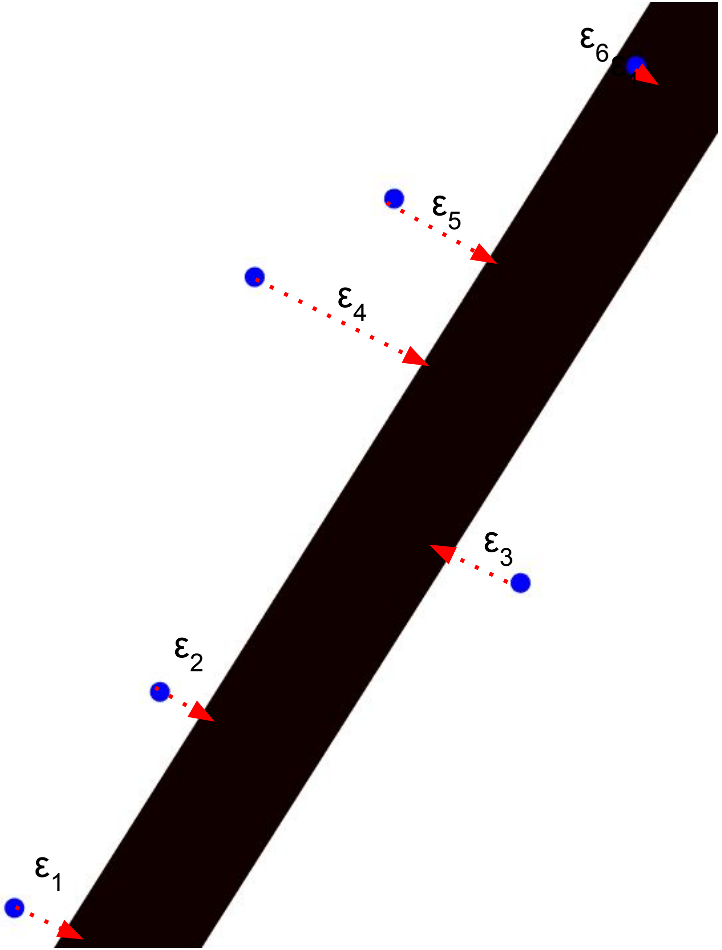
Hipótesis para regresión lineal univariada

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

Hipótesis para regresión lineal multivariada

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

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



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

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

¿Cómo definimos ε?

MSE

$$\varepsilon_i = (y_i - h(x_i))^2 = y_i - (x_{i1} + x_{i2} + b))^2$$

MAE

$$\varepsilon_i = |y_i - h(x_i)| = |y_i - (x_{i1} + x_{i2} + b)|$$

¿Cómo definimos la función de pérdida?

MSE

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

MAE

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

¿Qué faltaría?

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

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

Cambiar parámetros → Hallar derivadas respecto a los parámetros

Hallar : $$\frac{\partial \mathcal{L}}{\partial w_1}, \frac{\partial \mathcal{L}}{\partial w_2}, \frac{\partial \mathcal{L}}{\partial b}$$

$$\frac{\partial \mathcal{L}}{\partial b} =$$

Hallar : $$\frac{\partial \mathcal{L}}{\partial w_1}, \frac{\partial \mathcal{L}}{\partial w_2}, \frac{\partial \mathcal{L}}{\partial b}$$

$$\frac{\partial \mathcal{L}}{\partial w_1} =$$

Hallar : $$\frac{\partial \mathcal{L}}{\partial w_1}, \frac{\partial \mathcal{L}}{\partial w_2}, \frac{\partial \mathcal{L}}{\partial b}$$

$$\frac{\partial \mathcal{L}}{\partial w_2} =$$

```r
fit <- lm(db$Employed~., db)
b = fit$coefficients[1]
w1 = fit$coefficients[2]
w2 = fit$coefficients[3]
X1 = db$Year
X2 = db$Population
Y = db$Employed
open3d()
plot3d(X1, X2, Y,  col = "blue",size =15, xlab =  "Year", ylab="Population", zlab = "Employed estimation")
planes3d(w1,w2,-1,b, col="red")
```

Video

Generalizando a un espacio k dimensional

Hipótesis

Generalizando a un espacio k dimensional

Función loss

Generalizando a un espacio k dimensional

Hallar las k + 1 derivadas

```python
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
```

3 Regresión Lineal multivariable


INGENIERÍA
MECATRÓNICA
BIOINGENIERÍA
CIENCIA DE
LA COMPUTACIÓN
UTEC
UNIVERSIDAD DE INGENIERÍA
Y TECNOLOGÍA