FÓRMULA REG. LIN

$$\underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{\text{coeficientes}} = \underbrace{(X^T \cdot X)}_{X}^{-1} \cdot X^T y$$

$X \rightarrow \begin{bmatrix} \boxed{\text{dato 1}} \\ \boxed{\text{dato 2}} \end{bmatrix}$

y: $\begin{bmatrix} \boxed{\ } \\ \\ \\ \end{bmatrix} \rightarrow y$ del dato 1

Ejemplo del otro día:

$$y = \beta_2 x_2 + \beta_1 \cdot x_1 + \beta_0 \cdot 1$$

si hay β_0, aquí se pone "1"

$$\begin{bmatrix} \beta_2 \\ \beta_1 \\ \beta_0 \end{bmatrix} \quad \left\{ \begin{array}{l} X = \begin{bmatrix} salario_0 & tam\_fam_0 & 1 \\ salario_1 & tam\_fam_1 & 1 \\ \ldots & \ldots & 1 \end{bmatrix} \\ \\ Y = \begin{bmatrix} m^2\_0 \\ m^2\_1 \\ \vdots \end{bmatrix} \end{array} \right.$$

EFECTOS INTERACCIÓN (ampliación reg. lin)

multicolinearidad $\longrightarrow$ dos o más variables nos aporten la misma inform.

$\Downarrow$

hay columnas en X que son lin. dep. (iguales)

$$X = \begin{array}{cc} \text{man} & \text{woman} \\ \left[\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right. & \left.\begin{array}{c} 0 \\ 1 \\ 1 \end{array}\right] \end{array}$$

<u>mult. perfecta</u>

→ misma info

$$X = \begin{bmatrix} 21 & 2000 \\ 34 & 1987 \\ 36 & 1985 \end{bmatrix}$$

$(X^T X)^{-1}$ → falla la inversa → falla el modelo

↓

no puedo calc los β

<u>mult. imperfecta</u> → no es relación perfecta pero casi

fam familia ~ comida cons. por familia

experiencia ~ edad

¿Cómo lo sabemos? $\longrightarrow$ matriz correlaciones

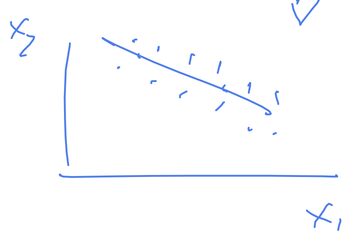
	x_0	x_1	x_2
x_0	1	-1	0.7
x_1	-1	1	0.6
x_2	0.7	0.6	1

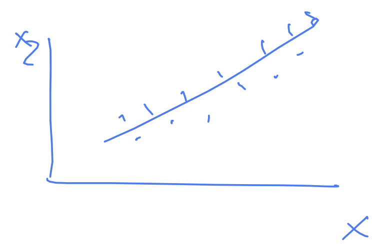
$x_0 \rightarrow \begin{cases} 0 & \text{no hombre} \\ 1 & \text{hombre} \end{cases}$

$x_1 \rightarrow \begin{cases} 0 & \text{no mujer} \\ 1 & \text{mujer} \end{cases}$

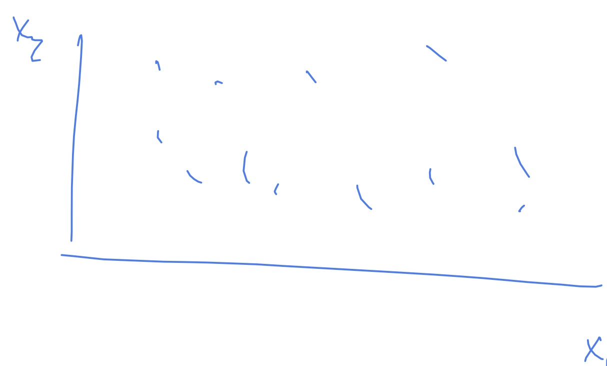
$x_2 \rightarrow$ peso

$$\rho = \frac{Cov(x_1, x_2)}{\sigma_{x_1} \cdot \sigma_{x_2}} = \text{entre } -1 \text{ y } 1$$





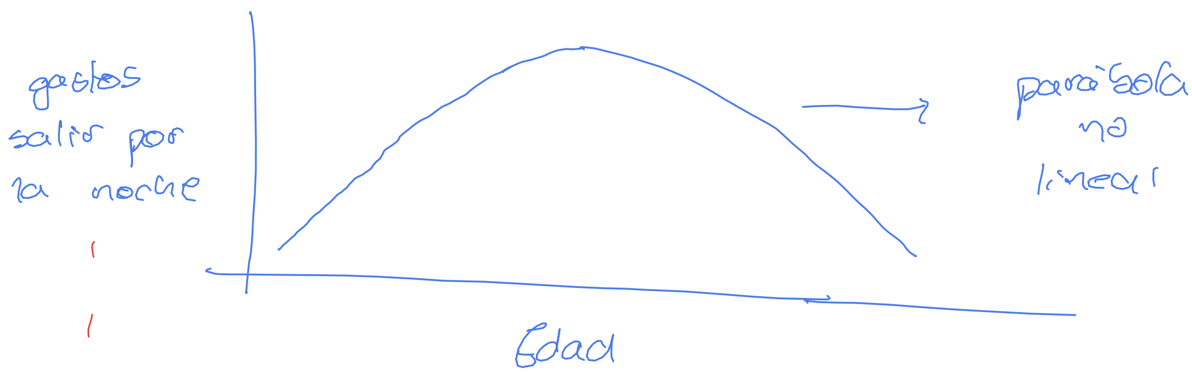
$\rho = 0 \longrightarrow$ no hay relación entre vars



efectos cuadráticos → relaciones de una variable con ella misma que captan o explican mejor el modelo

$$Y \leftarrow \underbrace{\beta x_2 + \beta_1 x_1}_{\text{modelo lineal}} \; \dots$$

¿Qué pasa cuando los datos no son lin?



$$y = \beta_2 \cdot \underbrace{x_1 \cdot x_1}_{x_1^2} + \beta_1 x_1 + \beta_0$$

además, podemos tener más var.

$$y = \underline{\beta_3 \cdot x_3} + \beta_2 \cdot x_1^2 + \beta_1 x_1 + \beta_0$$

efectos interacción → efecto cuadrático PERO entre dos vars. distintas

nº copas

0 hombre / 1 mujer ⇒ β_1 con las mismas copas → cuánto más si eres mujer

$$\text{nivel\_alcohol} = \beta_2 x_2 + \beta_1 \cdot x_1 + \beta_0$$

1 hombre 0 mujer

ejemplo interpr. coef.

$$17.23 = \overbrace{5.4}^{\beta_2} \cdot 3 + \overbrace{1.4}^{\beta_1} \cdot 0 + 1.03$$

$$18.63 = 5.4 \cdot 3 + 1.4 \cdot 1 + 1.03$$

para las mismas copas, una mujer tiene 1.4 más de nivel de alcohol.

$$\text{nivel\_alcohol} = \beta_3 \overbrace{x_2 \cdot x_1}^{\text{interacc.}} + 5.4 \cdot x_2 + \beta_1 x_1 + \beta_0$$

||| si soy hombre → 1 copa más ⇒ 5.4 sube niv_alc

si soy mujer → " " " ⇒ 5.4 " "

puede no ser verdad

$x_2 \cdot x_1 \longrightarrow$ arregla esto

$x_2 \cdot x_1 = 0$ si soy hombre

$x_2 \cdot x_1 \neq 0$

$= x_2$

si soy mujer

$\downarrow$

$x_1 = 1$

β_3: cuánto más me afecta una copa más si soy mujer

si $\Delta x_2 = 1 \xrightarrow{H} \Delta\text{nivel\_alch} = \overbrace{5.4}^{\beta_2} + 0$

(si tomo 1 copa más) $\xrightarrow{M} \Delta\text{nivel\_alch} = \overbrace{5.4}^{\beta_2} + \beta_3$

BIAS VARIANCE

X → diana (OK)

O → estimación

Low Variance

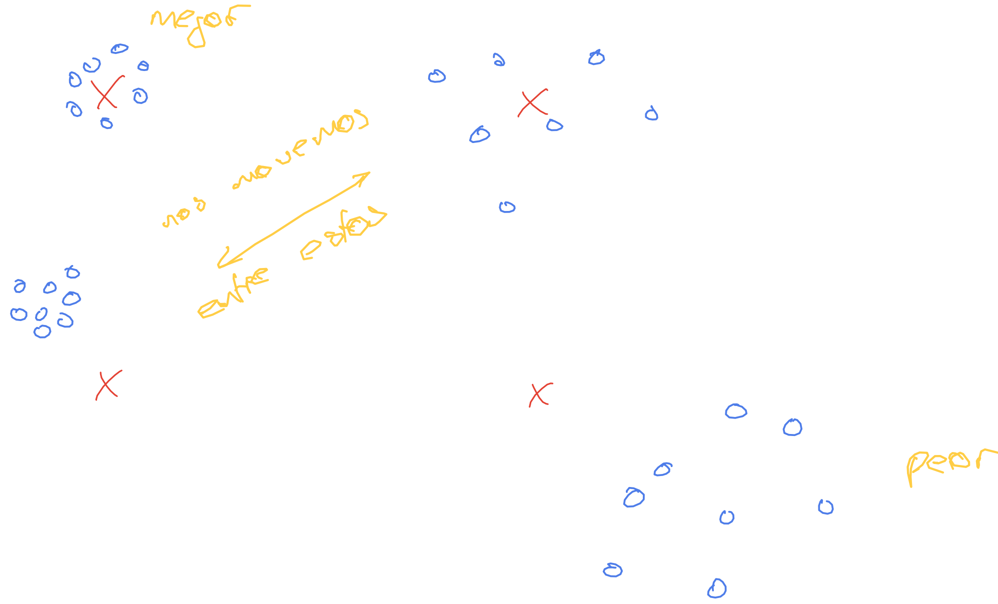
High variance ⟷ overfitting

Low Bias

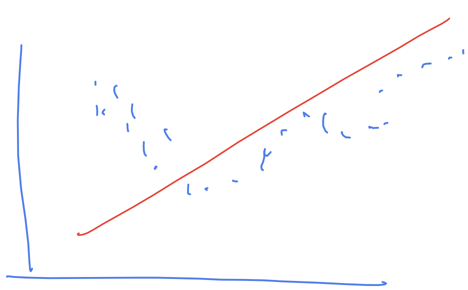
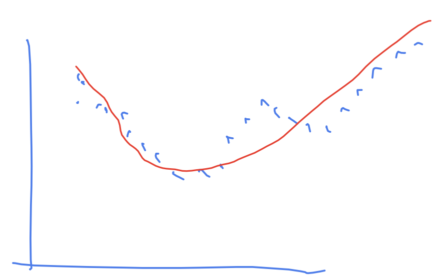
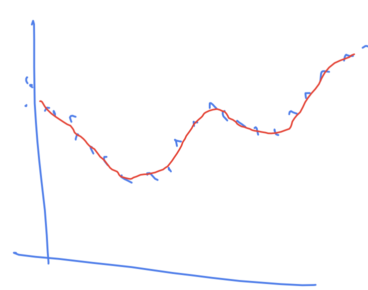
High Bias

↓

tenemos una variable irrelevante



Modelos pueden ser

Underfitting	Balanced/generalizar	Overfitting
		
te falta info por capturar	↑ mejor	explica demasiado