Dado $X_1, \dots, X_n$ un conjunto de mediciones e $Y_1, \dots, Y_n$ conjunto de respuestas. Vamos a llamar a X variable regresoras / independientes y a Y la llamaremos variable de respuesta / variable dependiente.

En general, vamos a asumir que existe una relación entre X e Y de la forma:

$$Y = f(X,\theta) + \varepsilon \leftarrow \text{ruido Gaussiano}$$

En Predicción la idea es inferir $\hat{Y} = \hat{f}(X)$, $\hat{f}$ es el estimador de f, $\hat{Y}$ es la estimación de Y. La precisión de la medición de Y tiene dos comp. de error: una componente irreducible (que es indep. de nuestros datos) y una comp. reducible.

Asumiendo que un X fijo y un $\hat{f}$ conocido vamos a calcular el error cuadrático medio entre Y y su predicción $\hat{Y}$.

$$\mathbb{E}(Y-\hat{Y})^2 = E(f(X)+\varepsilon-\hat{f}(X))^2$$

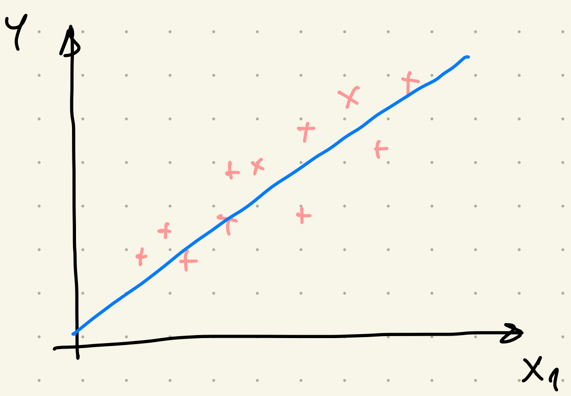
$$= \underbrace{E(f(X)-\hat{f}(X))^2}_{\text{componente reducible}} + E(\varepsilon^2) \leftarrow \overbrace{Var(\varepsilon)}^{\text{componente irreducible}}$$

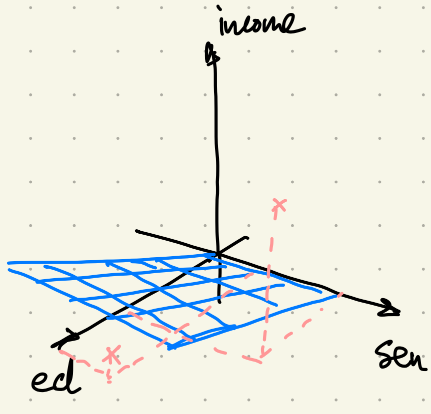
La f más sencilla es pensar una relación lineal → f es super simple
↳ f es fácil de interpretar
↳ no son muy precisos

Considerando un modelo paramétrico, vamos a pensar a f con la forma:

$$f(X) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i$$

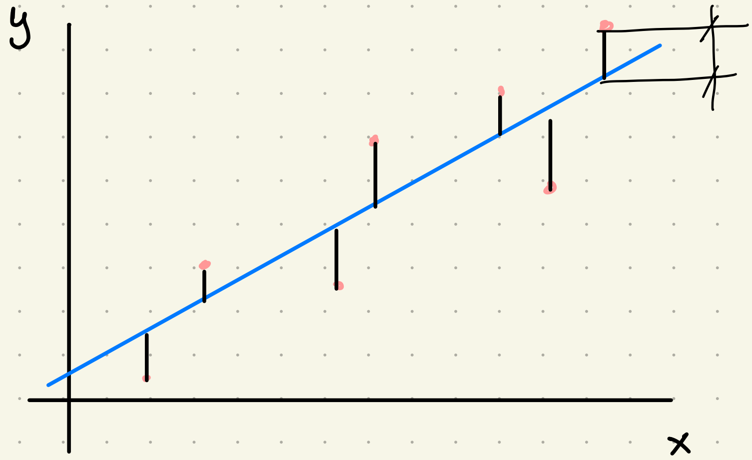
$$\text{income} \approx \beta_1 \cdot \text{education} + \beta_2 \, \text{seniority}$$





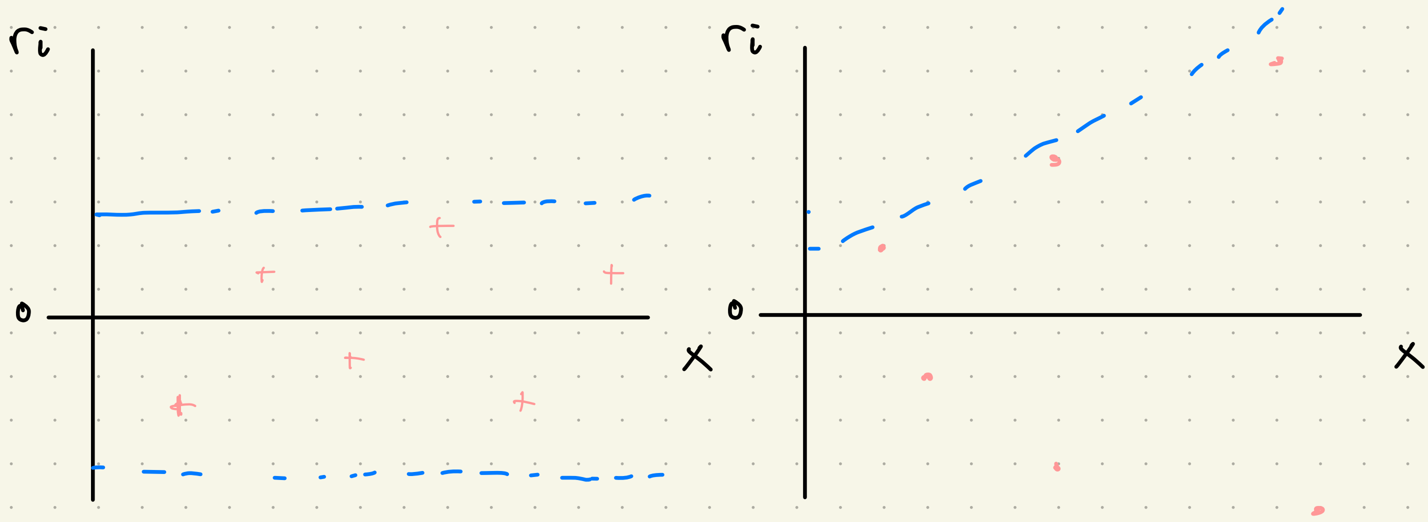
Condiciones del modelo:

El modelo lineal supone, los regresores sean independientes, los errores iid (independientes e identicamente distribuidos), $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$. (homocedasticidad). $\varepsilon \sim N(0, \sigma^2)$.



$y_i - \hat{y}_i = r_i$

↑ residuo



residuo homocedastico　　　　heterocedastico

Con estas cond. estoy limitando como encontrar $\hat{f}$ como el mejor

Tenemos 3 met
- → MSE (mean square error)
- → ML (Maximum Likelihood)
- → MAP (maximum a posteriori)

MSE: partimos de un dataset $D = \{ (x_i, y_i) \ \forall i \in [1, \dots, n], \ x_i \in \mathbb{R}^{m \times 1} \}$

$$\mathcal{E}(\beta) = \sum_{n=1}^{N} (y_n - \hat{f}(x_n))^2 = \sum_{n}^{N} \left(y_n - \beta_0 - \sum_{i}^{m} x_i \cdot \beta_i\right)^2$$

al vector X_i le agrego 1 para representar el termino indep. (β_0), entonces $x_i = [1, x_1, \dots, x_n]$, con esto, puedo escribir:

$$\mathcal{E}(\beta) = \sum_n \left(y_n - \sum_j \beta_j \cdot x_j\right)^2$$

$$= (y - X\beta)^t (y - X\beta)$$

Ahora tenemos $\mathcal{E}(\beta)$ y quiero optimizarlo $\Rightarrow$ tengo que minimizarlo:

$$\partial_\beta \mathcal{E} = 0 = \partial_\beta \left[(y - X\beta)^t (y - X\beta)\right]$$

$$= -2X^t (y - X\beta)$$

$$= X^t (y - X\beta)$$

$$= X^t y - X^t X \beta \Rightarrow \boxed{\hat{\beta} = \underbrace{(X^t X)^{-1} X^t}_{H} y}$$

$\hat{y} = H y \rightarrow H = X (X^t X)^{-1} X^t$ (pseudo inversa Moore Penrose)

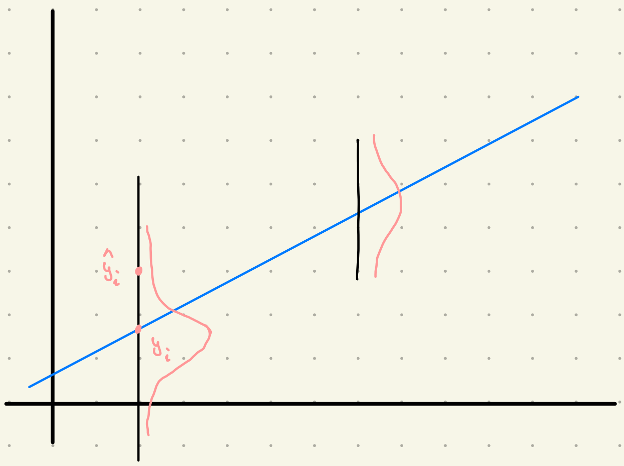
2. Método ML (Maximum Likelihood - Máxima verosimilitud):

$\hat{\beta}_{ML}$, partimos de $y_i = \hat{f}(x_i) + \varepsilon \quad \varepsilon \sim N(0, \sigma^2)$

$$P(Y \mid X, \beta) = \mathbb{P}(y_1, y_2, \dots, y_n \mid x_1, \dots, x_n, \beta)$$

x sup. iid

$$= \prod_{n=1}^{N} \mathbb{P}(y_n \mid x_n, \beta) \sim \mathcal{N}(y_n / x_n^t \theta, \sigma^2)$$



buscamos el $\hat{\beta}_{ML} = \arg\max \left(P(Y \mid X, \beta)\right)$:

$$\mathcal{L}(\beta) = \arg\max \left(\underbrace{P(y \mid x, \beta)}_{\prod \ldots} \right)$$

$$\log\left(\mathcal{L}(\beta)\right) = \arg\max\left(\log\left(P(y \mid x, \beta)\right)\right)$$

$$\mathcal{L}(\beta) = -\arg\min\left(\underbrace{\log\left(P(y \mid x, \beta)\right)}_{1}\right)$$

$$1 \quad \log\left(P(y_n \mid x_n, \beta)\right) = \frac{1}{2\sigma^2}\left(y_n - x_n^t \beta\right)^2 + c$$

$$\mathcal{L}(\beta) = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \left(y_n - x_n^t \beta\right)^2 = \frac{1}{2\sigma^2} (y - X\beta)^t (y - X\beta)$$

$$= \frac{1}{2\sigma^2} \| y - X\beta \|^2$$

Vamos a optimizar $\mathcal{L}(\beta)$:

$$\partial_\beta \mathcal{L}(\beta) = 0 = \partial_\beta \left(\frac{1}{2\sigma^2} (y - X\beta)^t (y - X\beta) \right)$$

$$= \partial_\beta \left(y^t y - 2 y^t X \beta_{ML} + \beta_{ML}^t X^t y \beta_{ML} \right)$$

$$= 0 - 2 y^t X + 2 \beta_{ML}^t X^t X$$

$$= - y^t X + \beta_{ML}^t X^t X$$

$$\hookrightarrow \quad \beta_{ML}^t X^t X = y^t X$$

$$\beta_{ML}^t = y^t X (X^t X)^{-1}$$

$$\boxed{\beta_{ML} = (X^t X)^{-1} X^t y}$$