

Regresión Lineal

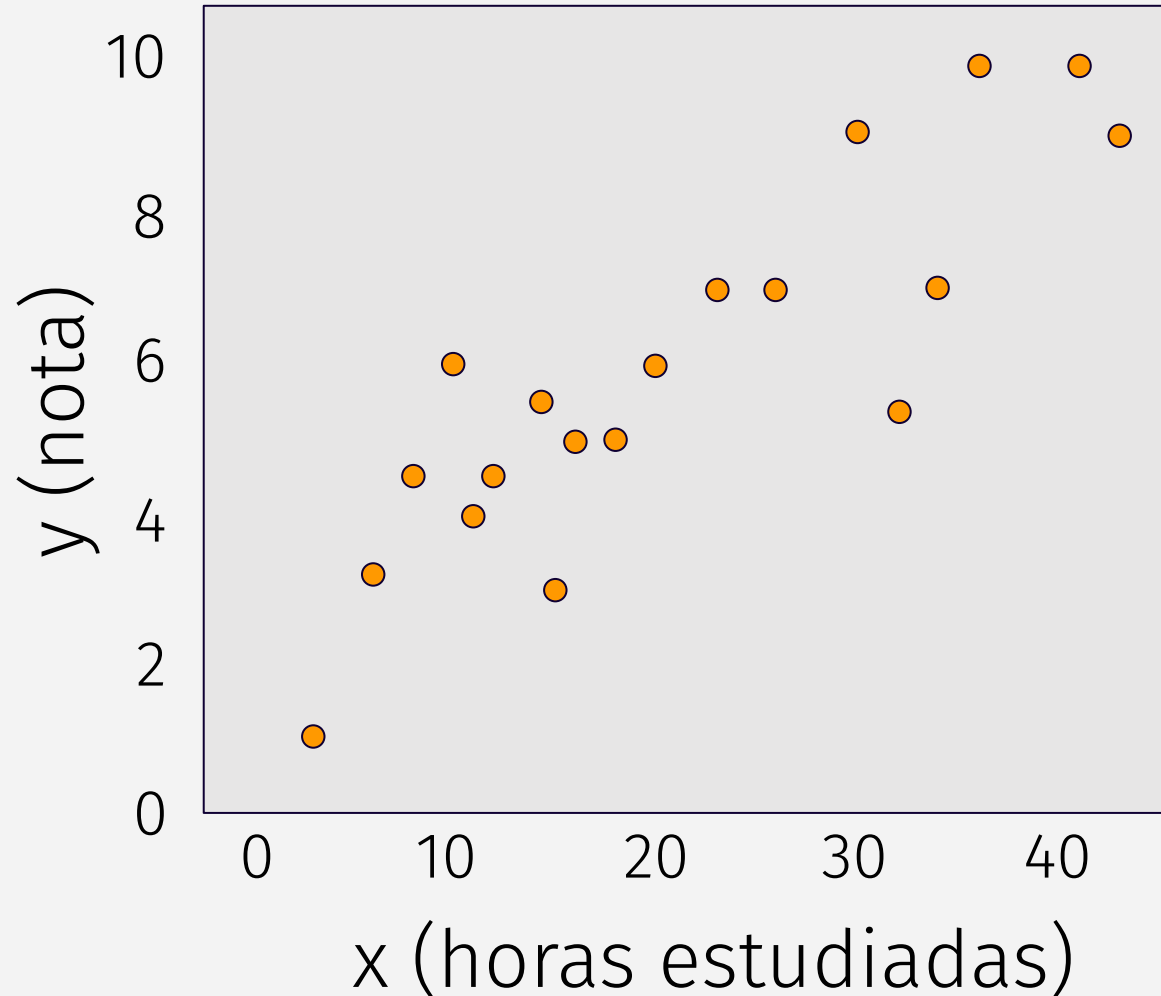
Introducción

Problema de Regresión

Ejemplos

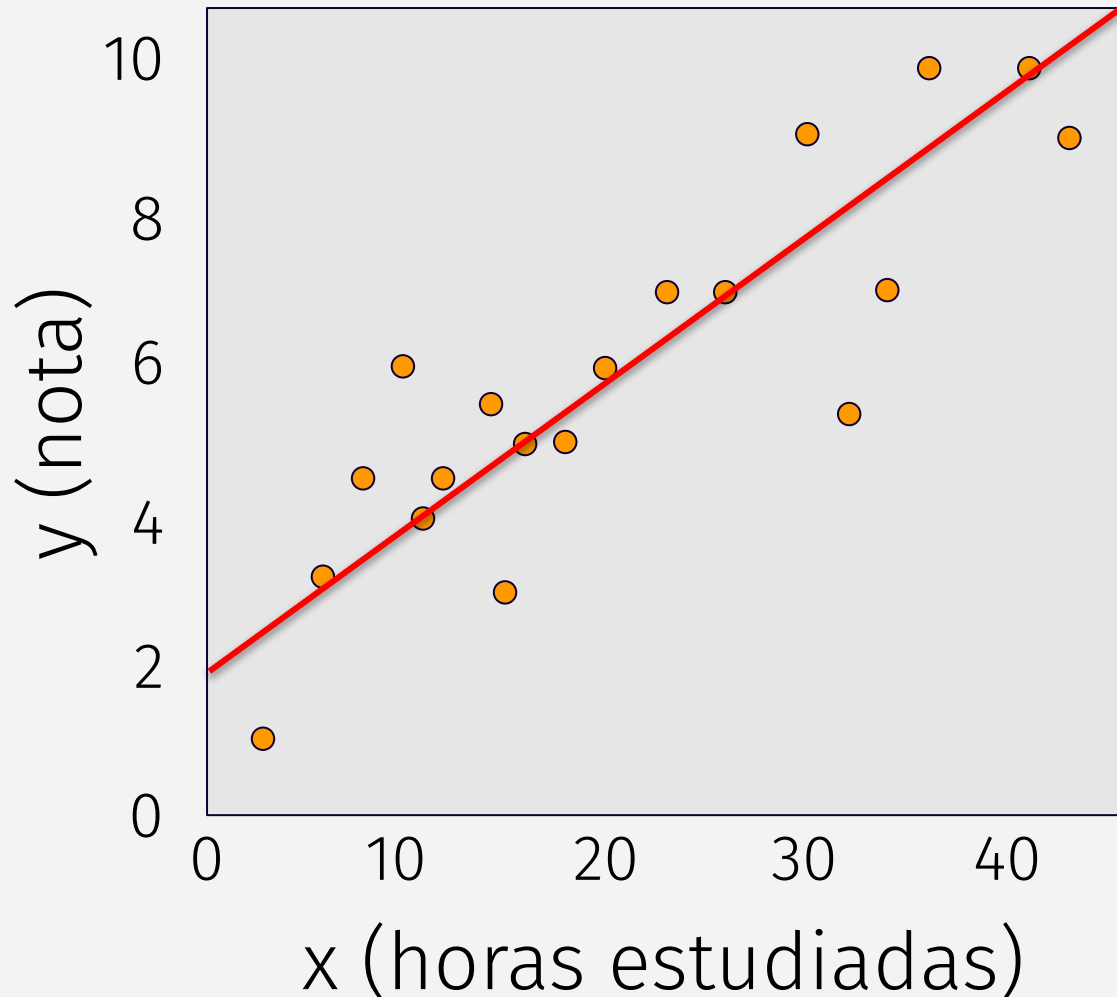
Horas	Nota
2	1
5	3.2
7	4.5
9	6
10	4
11	4.5
13.4	5.5
14	3
15	5

Visualización



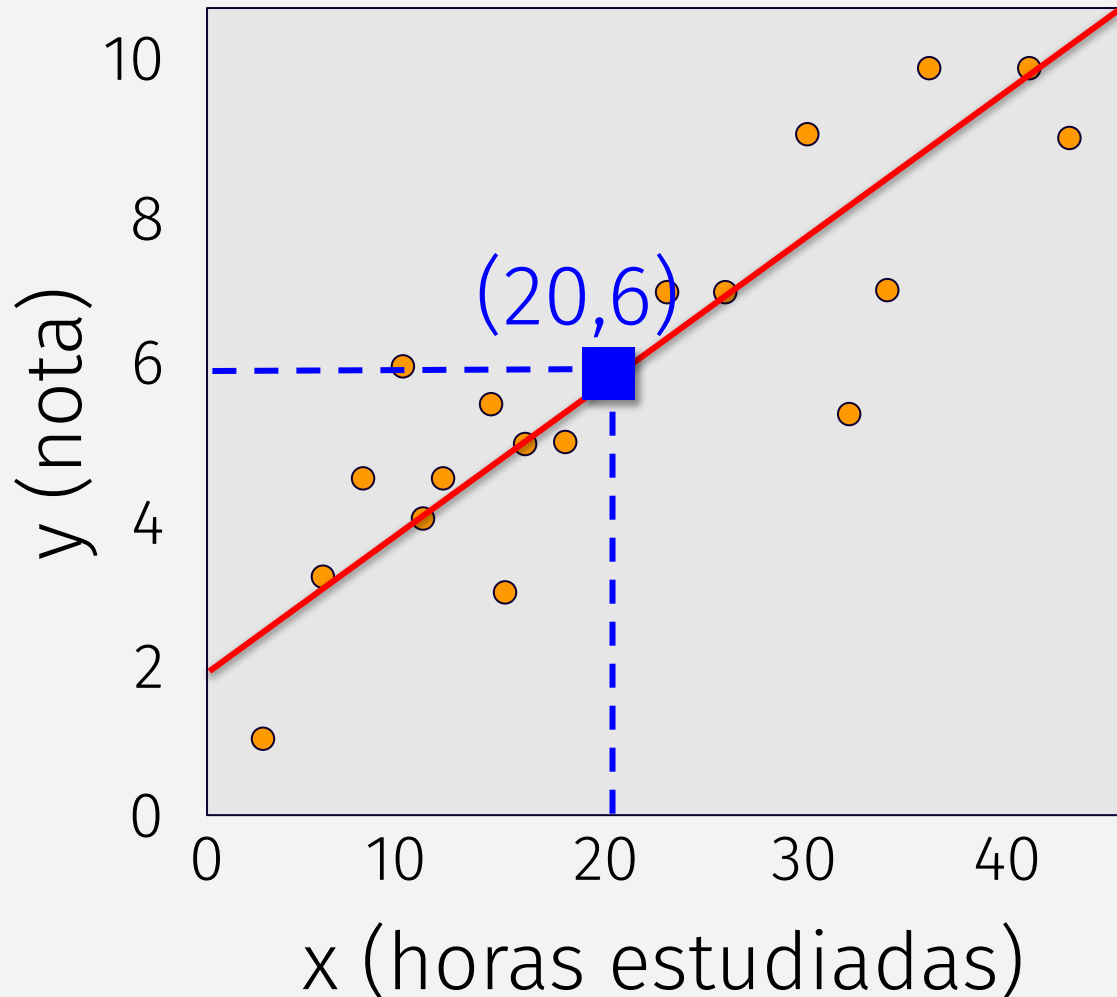
Si un nuevo alumno estudió $x = 20hs$, ¿cuál será su nota?

Modelo de Regresión Lineal (RL)



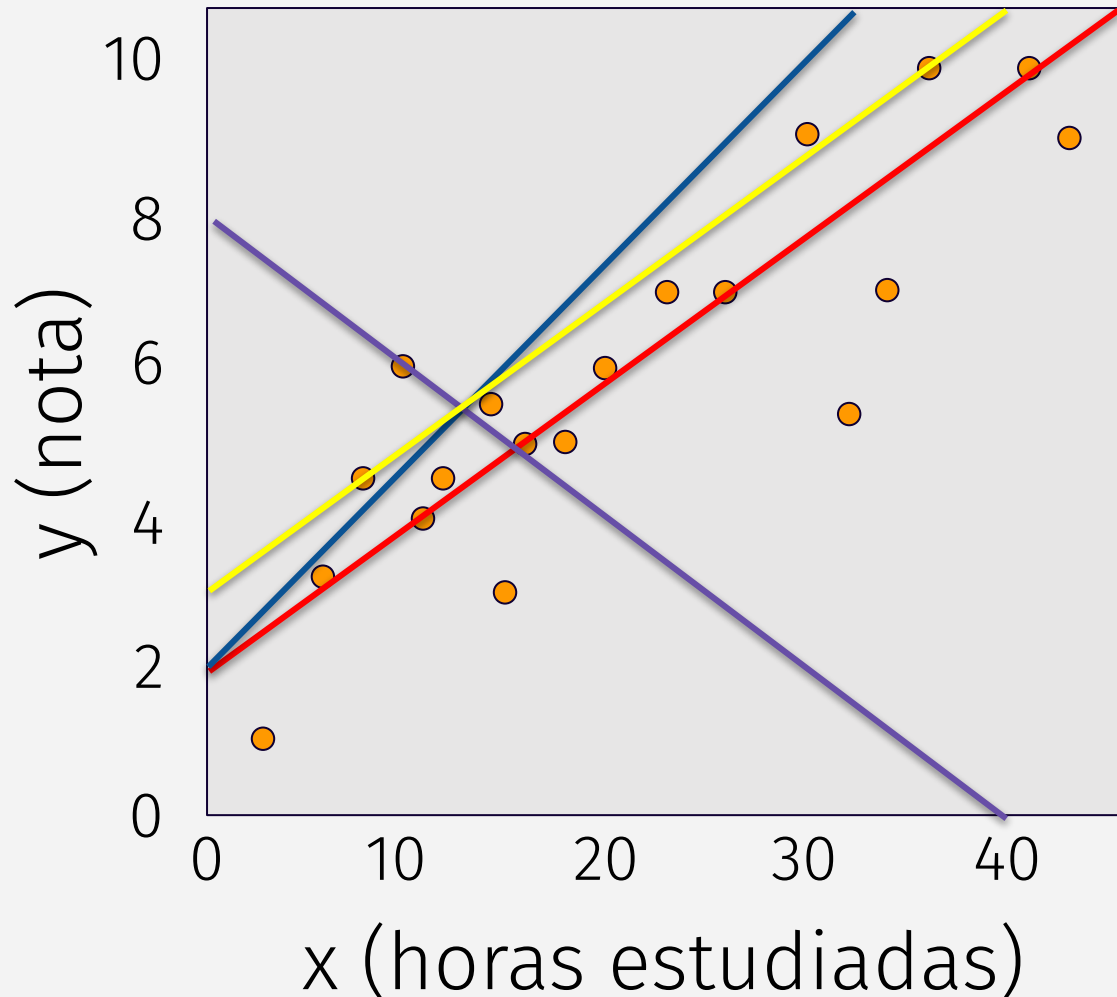
- Asunción
 - Relación $x - y$ es lineal
 - Modelo más simple
- Modelo
 - $y = m \times x + b$
 - 0
 - $f(x) = m \times x + b$

¿Cómo predecir con modelo de Regresión Lineal?



- Supongamos
 - $m = 0.5$
 - $b = 2$
 - $f(x) = 0,2 \times + 2$
- ¿Qué nota **predice** el modelo si **$x=20$** ?
 - $f(20) = m \times 20 + b$
 - $= 0,2 \times 20 + 2$
 - $= 6$
- Predice **$y=nota=6$**

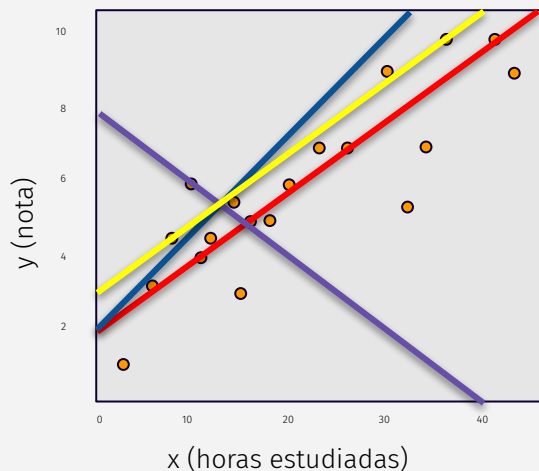
Valores de **m** y **b**



- Valores de **m** y **b** definen la recta
 - **m** = 0.20, **b** = 2
 - **m** = 0.20, **b** = 3
 - **m** = 0.26, **b** = 2
 - **m** = -0.20, **b** = 8
- **parámetros** del modelo
 - **m** indica la pendiente
 - **b** la ordenada a la origen

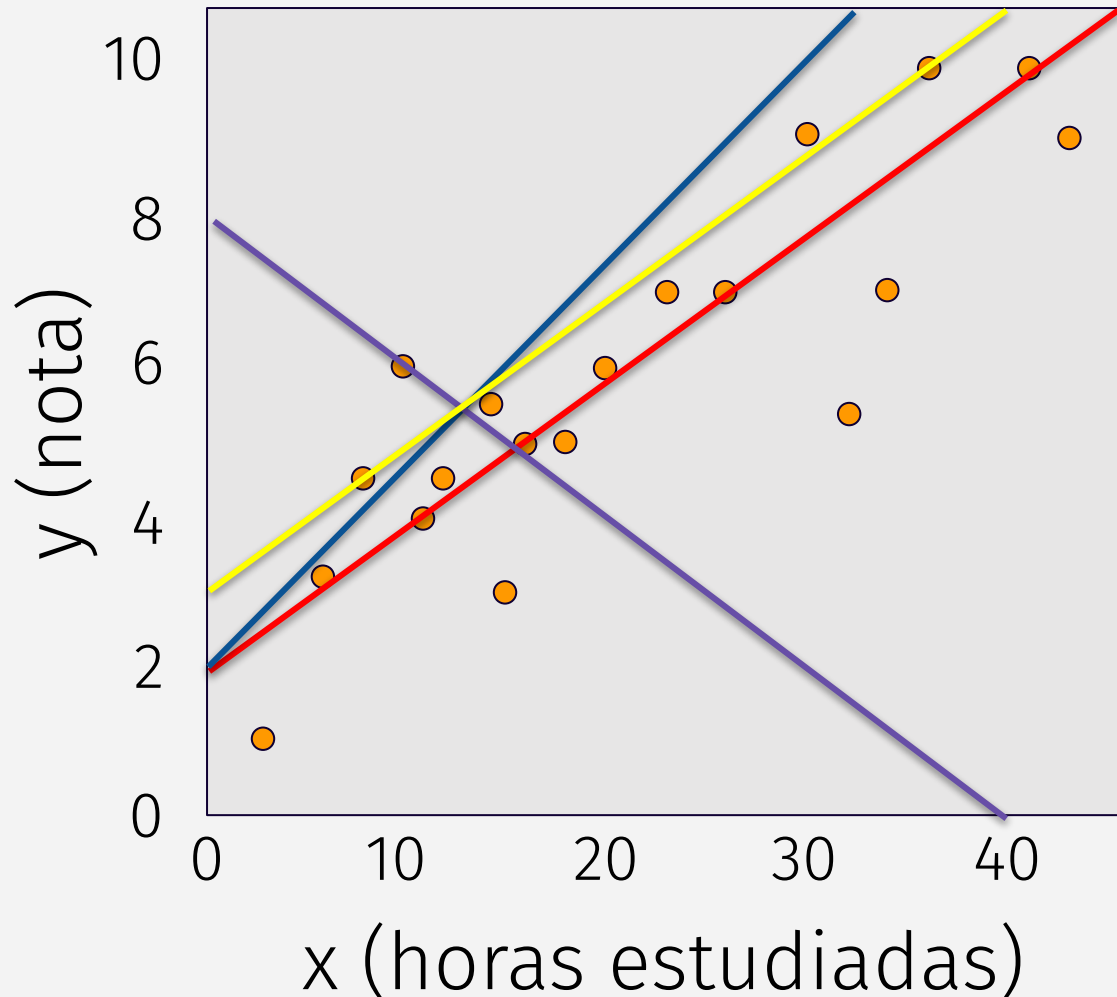
Valores de **m** y **b**

- Valores de **m** y **b** definen la recta
 - - **m** = 0.20, **b** = 2
 - - **m** = 0.20, **b** = 3
 - - **m** = 0.26, **b** = 2
 - - **m** = -0.20, **b** = 8



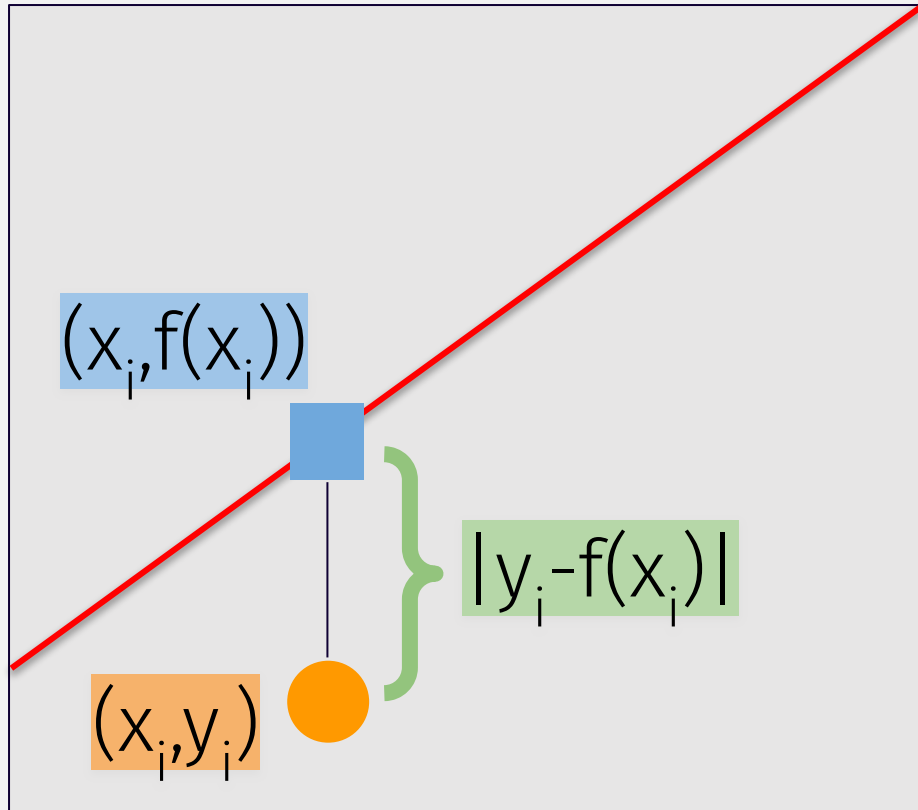
- Valores de **m** y **b** definen la recta
 - - $f(x) = 0.20 \times x + 2$
 - - $f(x) = 0.20 \times x + 3$
 - - $f(x) = 0.26 \times x + 2$
 - - $f(x) = 0.20 \times x + 8$
- Cada recta es un modelo **distinto**
- Misma familia de modelos
 - **Regresión Lineal**

Elección de **m** y **b**. Función de **error**



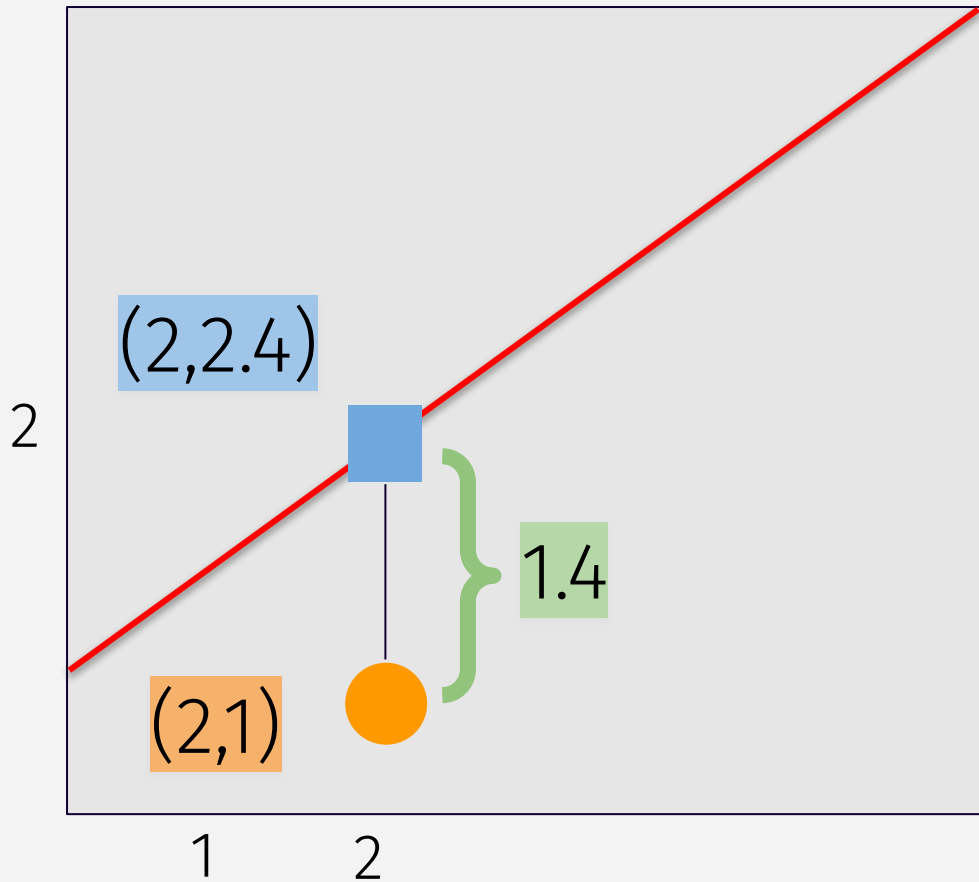
- Dado un conjunto de ejemplos
 - ¿Cómo elegir **m** y **b**?
 - Función de **puntaje** para cada **m** y **b**
 - Función de **error** **$E(m,b)$**
 - Ejemplo
 - $E(-) = 1.62$
 - $E(-) = 1.72$
 - $E(-) = 5.12$
 - $E(-) = 1.44$

$E_i(m,b)$: Error cuadrático del modelo para el ejemplo i



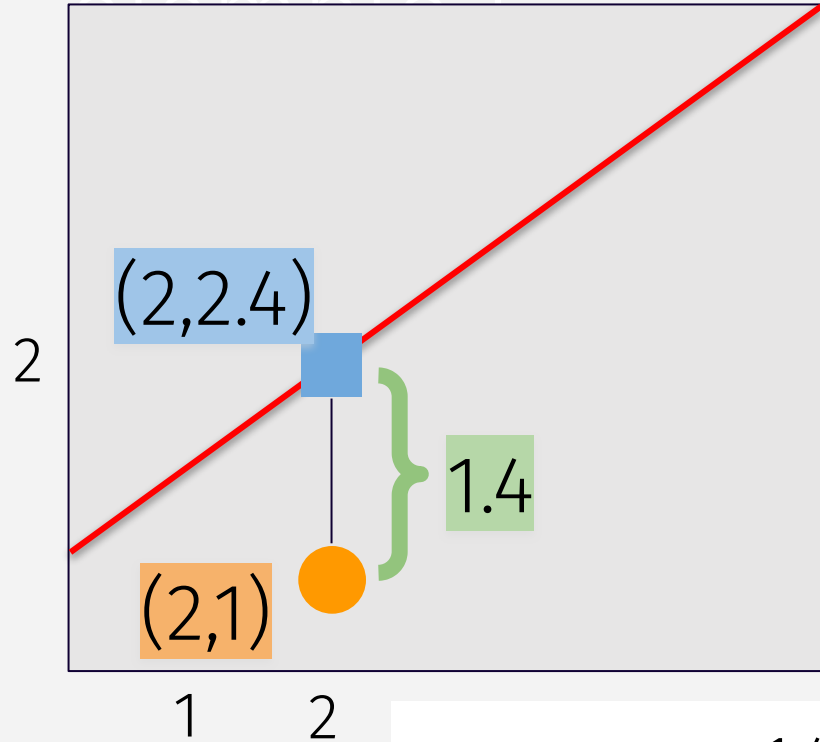
- Primero, definimos el **error** para **un solo ejemplo**
- $E_i(m,b)$
 - Error del ejemplo i para m y b
 - Distancia cuadrática entre
 - y_i : valor esperado
 - $f(x_i)$: valor predicho
- $E_i(m,b) = (y_i - f(x_i))^2$
 $= (y_i - m x_i + b)^2$

$E_i(m,b)$: Error cuadrático del modelo para el ejemplo i



- $E_i = (y_i - f(x_i))^2$
- Ejemplo
 - Estudió 2 horas, nota 1
 - $x_i = 2$
 - $y_i = 1$
 - $f(x_i) = f(2) = 0.2 \times 2 + 2$
 $= 2.4$
 - $y_i - f(x_i) = 1 - 2.4 = -1.4$
 - $(y_i - f(x_i))^2 = (-1.4)^2 = 1.96$
 - $E_i = 1.96$

$E_i(m,b)$: Error cuadrático del modelo para el



$$y_i - f(x_i) = -1.4$$

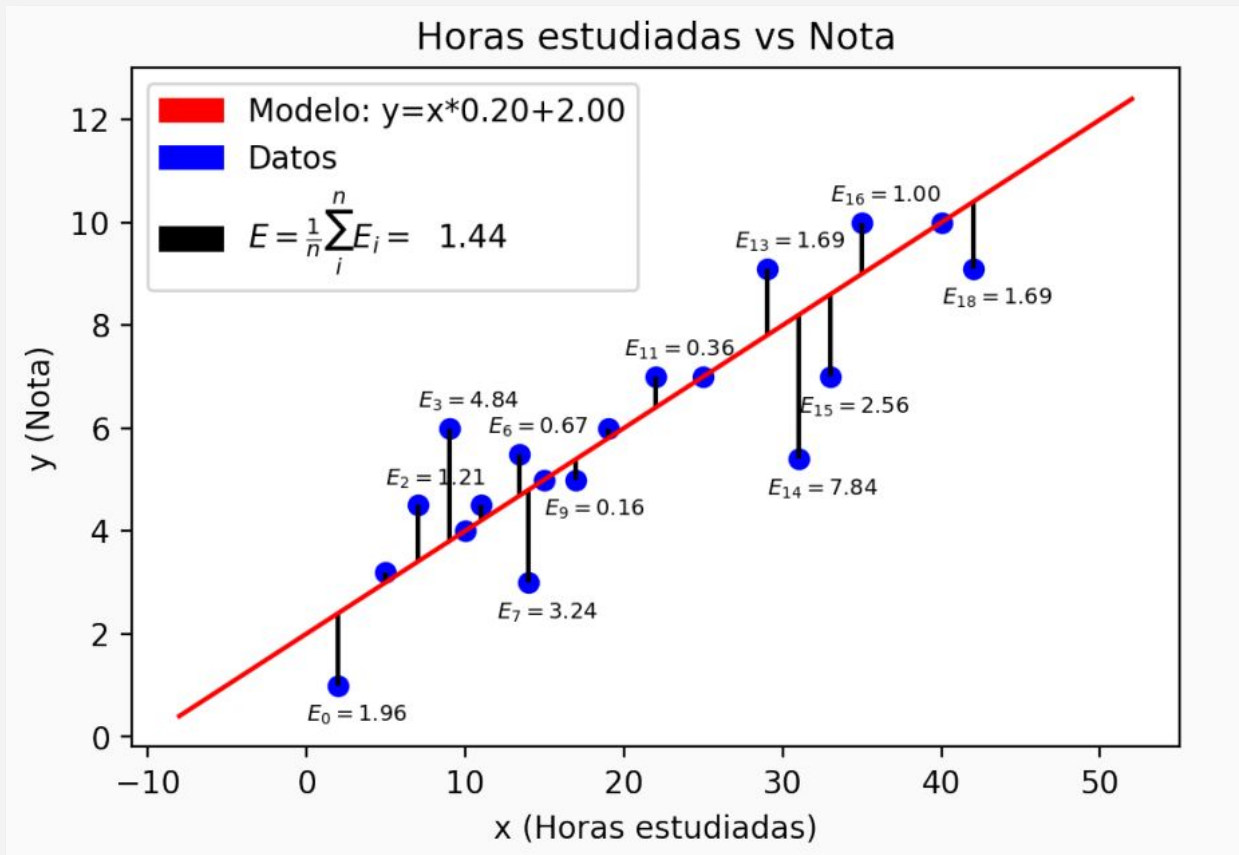
$$|y_i - f(x_i)| = 1.4$$

$$(y_i - f(x_i))^2 = 1.96$$

$$(y_i - f(x_i))^4 = 3.84$$

- $E_i = (y_i - f(x_i))^2$
 - ¿Por qué esta función de error?
- ¿Por qué no usar $y_i - f(x_i)$?
 - Valores negativos
- ¿Por qué no usar $|y_i - f(x_i)|$?
 - Valor absoluto
 - No es una función derivable
 - Difícil de optimizar
- ¿Qué efecto tiene el **2**?
 - Penaliza más errores grandes
 - $0.5^2 = 0.25$, $1^2 = 1$, $5^2 = 25$
- ¿Por qué no usar $(y_i - f(x_i))^4$?
 - Posible, pero penalizaría demasiado

$E(m,b)$: Error cuadrático medio de los ejemplos



- $E(m,b)$: Error cuadrático medio
 - Función de error **natural** para RL
 - Promedio de los E_i
- $E(m,b) = (1/n) \sum_i^n E_i(m,b)$

Ejemplo de cálculo de **E** para **m=0.2** y **b=2**

$$m = 0.2$$

$$b = 2$$

estudio nota

2

1

10

4

14

5

30

9

40

10

$$f(x_i)$$

$$f(x_i) - y_i$$

$$E_i$$

$$0.2 * 2 + 2 = 2.4$$

$$2.4 - 1 = 1.4$$

$$(1.4)^2 = 1.96$$

$$0.2 * 10 + 2 = 4$$

$$4 - 4 = 0$$

$$(0)^2 = 0$$

$$0.2 * 14 + 2 = 4.8$$

$$4.8 - 5 = -0.2$$

$$(-0.2)^2 = 0.04$$

$$0.2 * 30 + 2 = 8$$

$$8 - 9 = -1$$

$$(-1)^2 = 1$$

$$0.2 * 40 + 2 = 10$$

$$10 - 10 = 0$$

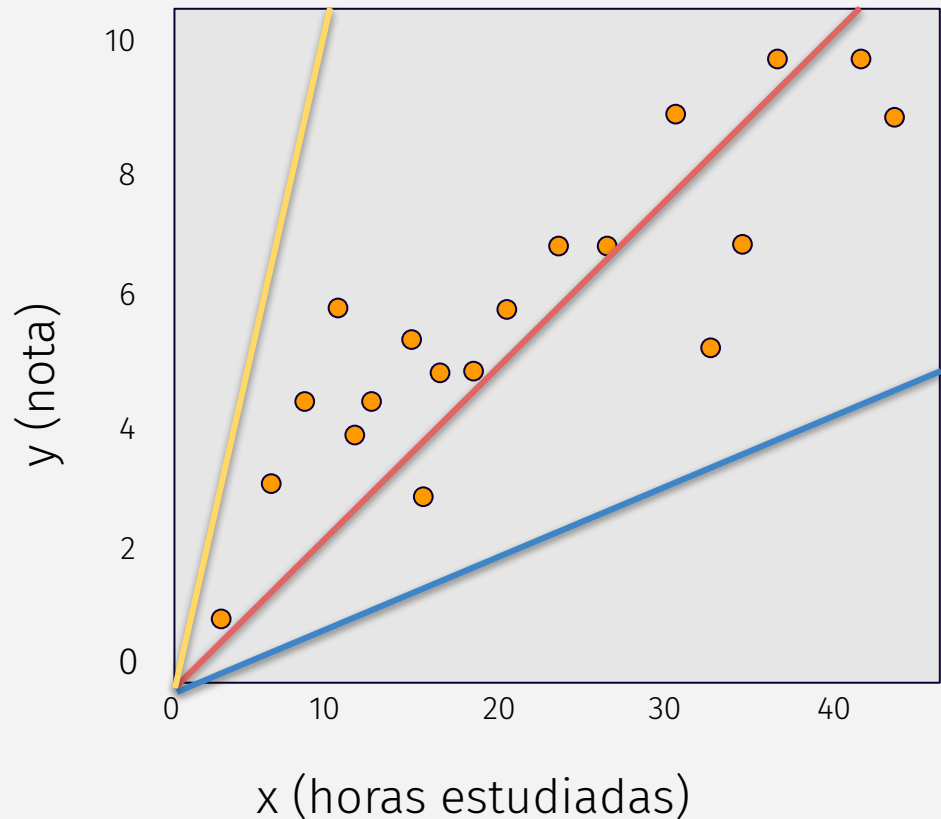
$$(0)^2 = 0$$

Error cuadrático medio

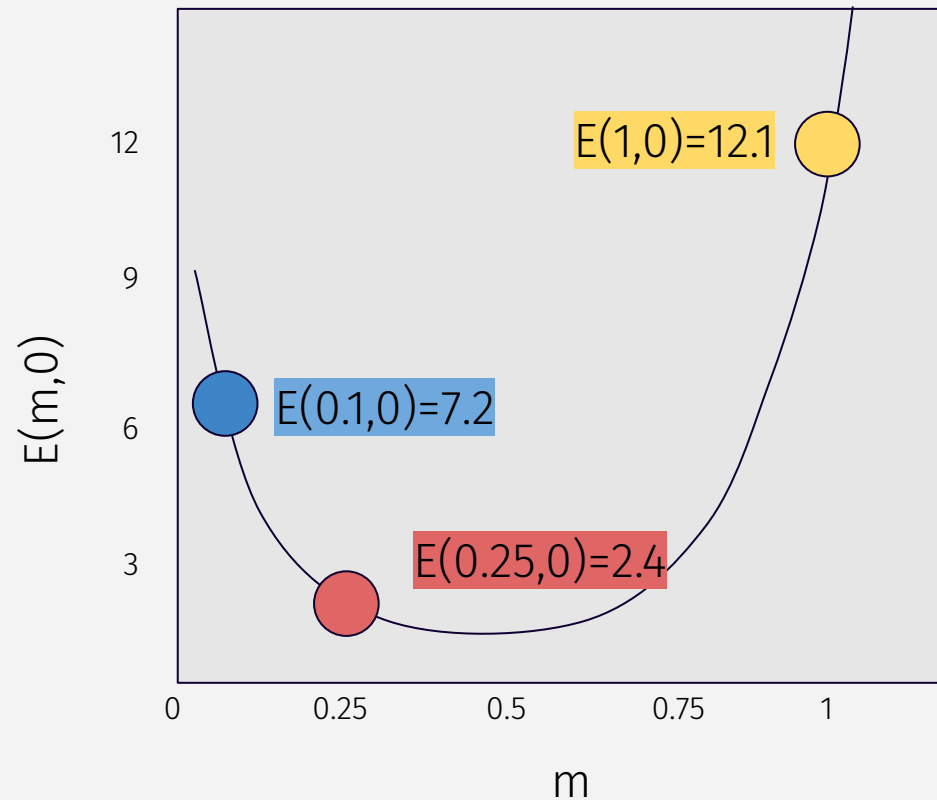
$$E = \frac{1}{n} \sum_i^n E_i = \frac{1.96 + 0 + 0.04 + 1 + 0}{5} = \frac{3}{5} = 0.6$$

Comprendiendo $E(m,b)$

- Asumamos $b=0$
 - Queda 1 parámetro: m
 - Probamos con $m=...$
 - 0.1 o 0.25 o 1

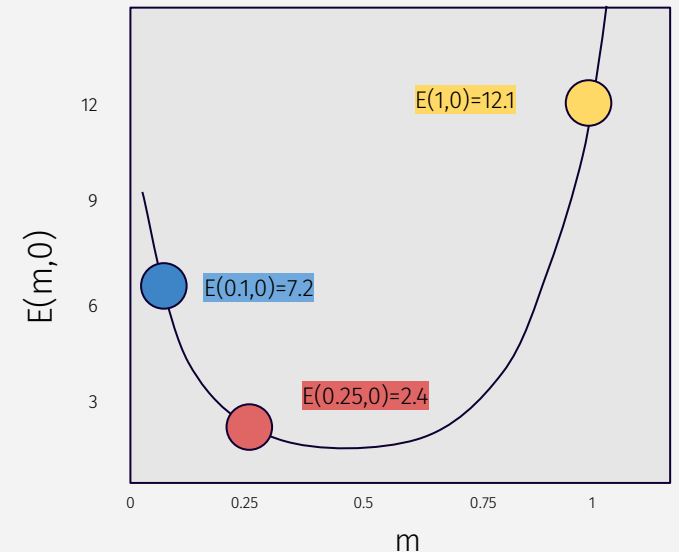
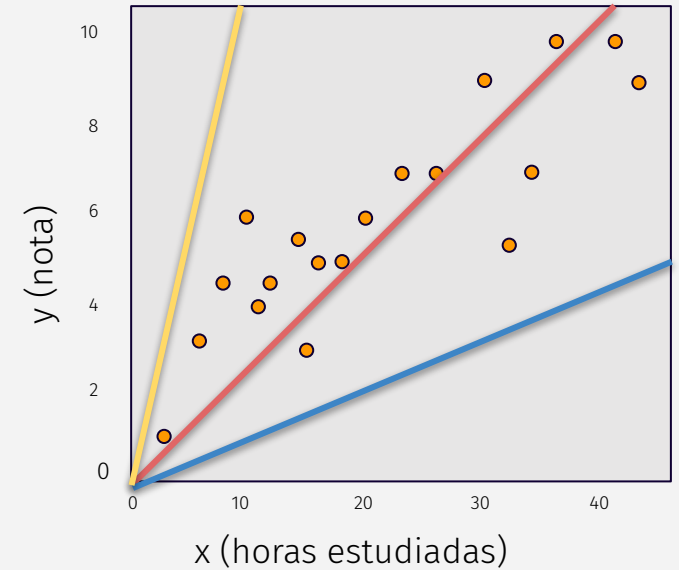


- Error $E(m,b) = E(m,0)$
 - 1D
 - Forma una parábola



Resumen

- Regresión Lineal
 - $f(x)=m x + b$
 - Modelo más simple
 - Asume relación lineal entre x e y
 - Aproximada
- Parámetros
 - **m** pendiente
 - **b** ordenada al origen
- Función de error **E**
 - Error cuadrático medio
 - Promedio de error de cada ejemplo E_i
 - Permite elegir valores óptimos de **m** y **b**



Ejercicio: Archivo **Regresión Lineal - Modelo.ipynb**

- **Probar**

- Modificar los parámetros **m** y **b**. Observar como cambia la recta y el error
- Modificar el valor de la variable **new_x**, observar los valores de $f(\mathbf{new_x})$ que predice el modelo (punto verde)

- **Interpretar**

- ¿Qué me dice el modelo respecto a la nota que obtienen los alumnos que no estudiaron nada?
- ¿Cuánto puedo esperar que aumente mi nota por cada hora de estudio?
- ¿Cuándo es más fácil la materia? (en relación a m)
- ¿Cuándo baja el error? ¿Puede ser 0?