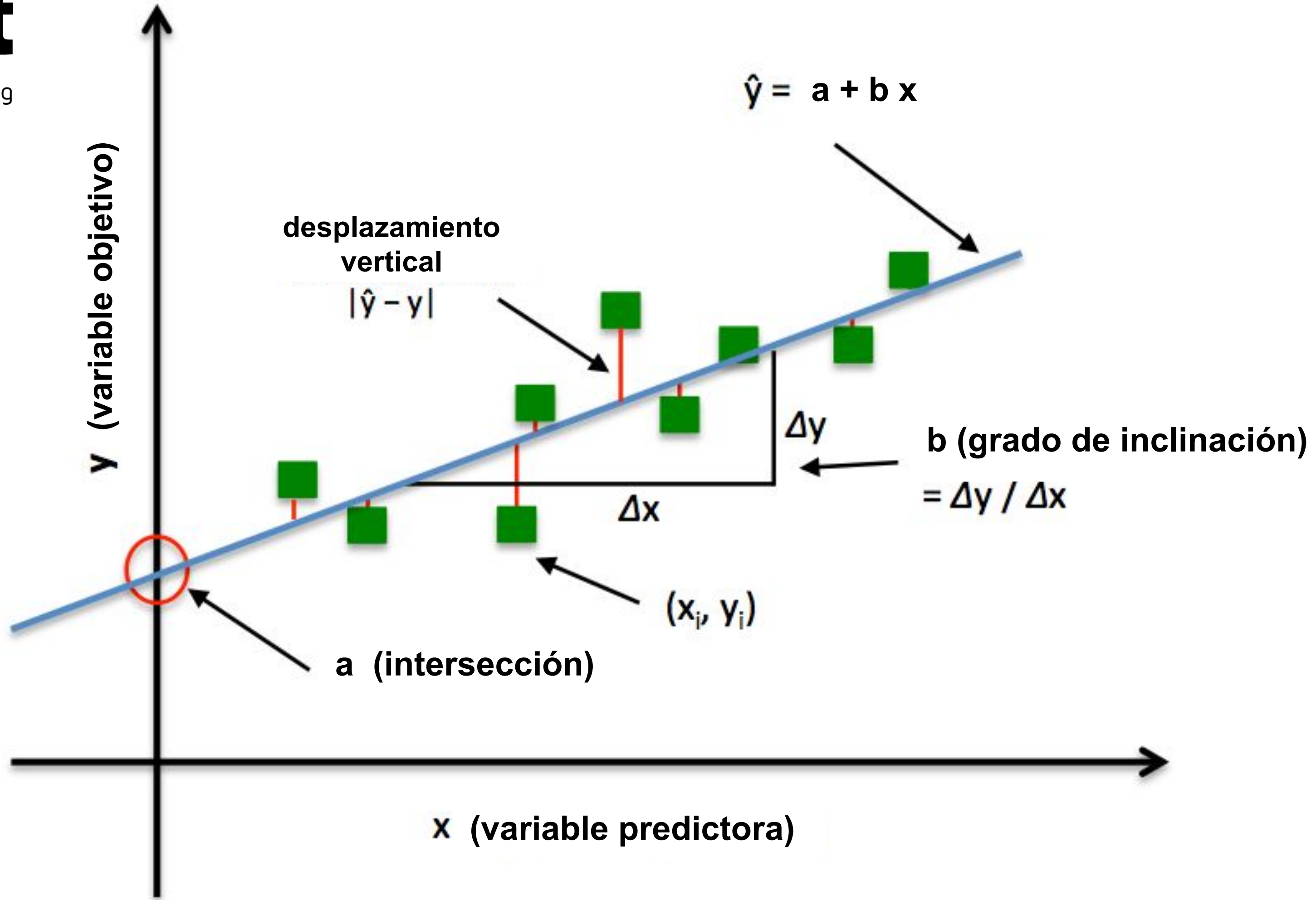
boost
accelerate your learning

Regresiones lineales y predicción de cantidades continuas : caso House Pricing

Sesión 4:

Pablo Benavides - Ricardo Obregón


ŷ = a + b x
desplazamiento vertical
|ŷ − y|
y (variable objetivo)
Δy
b (grado de inclinación)
= Δy / Δx
Δx
(xi, yi)
a (intersección)
x (variable predictora)

Hipótesis: $h(x) = a + b\,x$

Parámetros: a, b

Función de coste: $$J(a, b) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_{a,b}(x_i) - y_i \right)^2$$

Objetivo: $$\underset{a,\,b}{\text{minimize}}\; J(a, b)$$

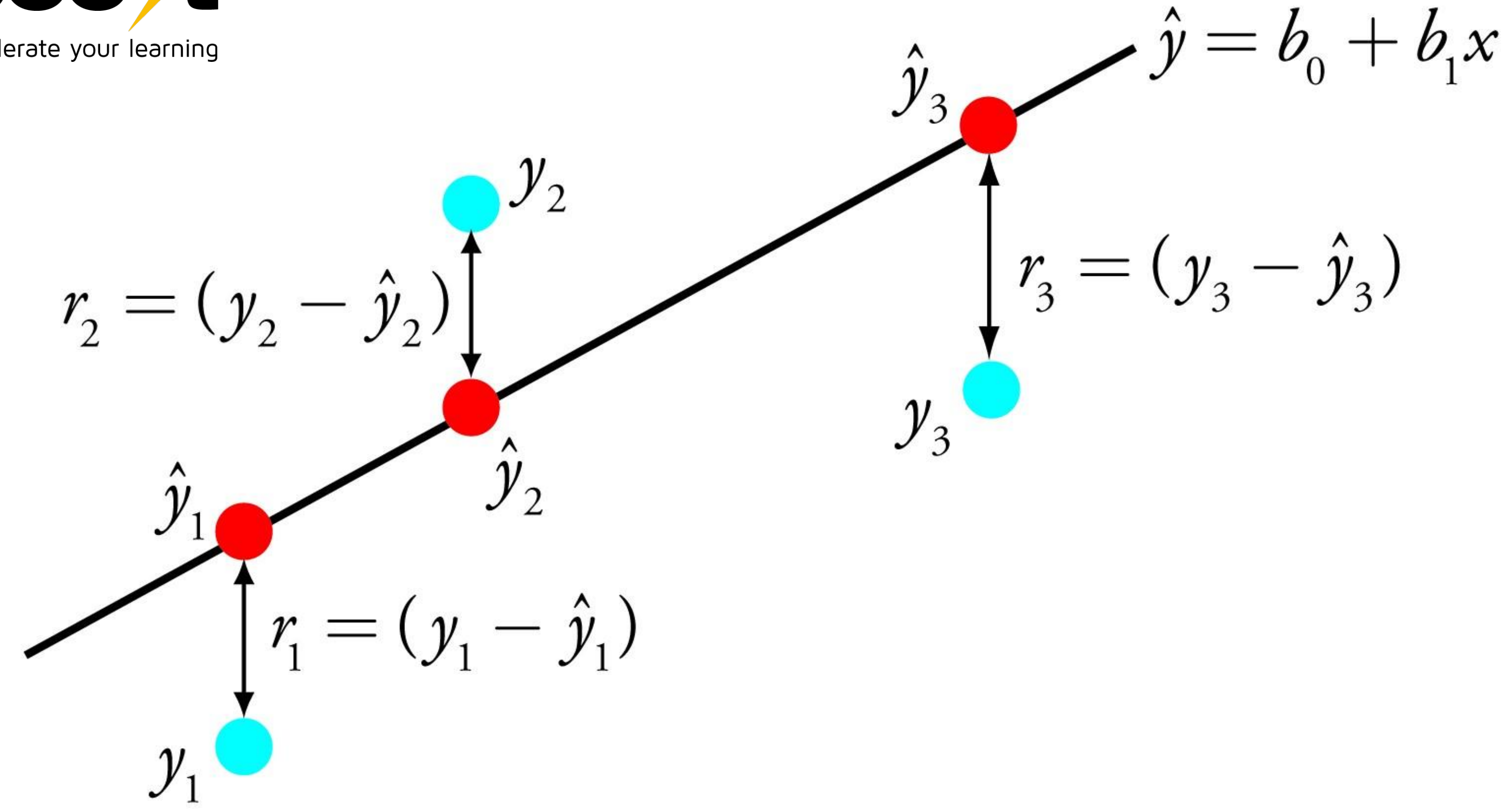
boost
accelerate your learning

$$Hipotesis : h(x) = a + bx$$

$$Parametros : a, b$$

$$Funcion\ de\ coste : J(a, b) = \frac{1}{2m}\sum_{i=1}^{m}(h_{a,b}(x_i) - y_i)^2$$

$$Objetivo : minimize_{a,b}\ J_{(a,b)}$$


$\hat{y} = b_0 + b_1 x$
$\hat{y}_3$
y_2
$r_3 = (y_3 - \hat{y}_3)$
$r_2 = (y_2 - \hat{y}_2)$
y_3
$\hat{y}_2$
$\hat{y}_1$
$r_1 = (y_1 - \hat{y}_1)$
y_1

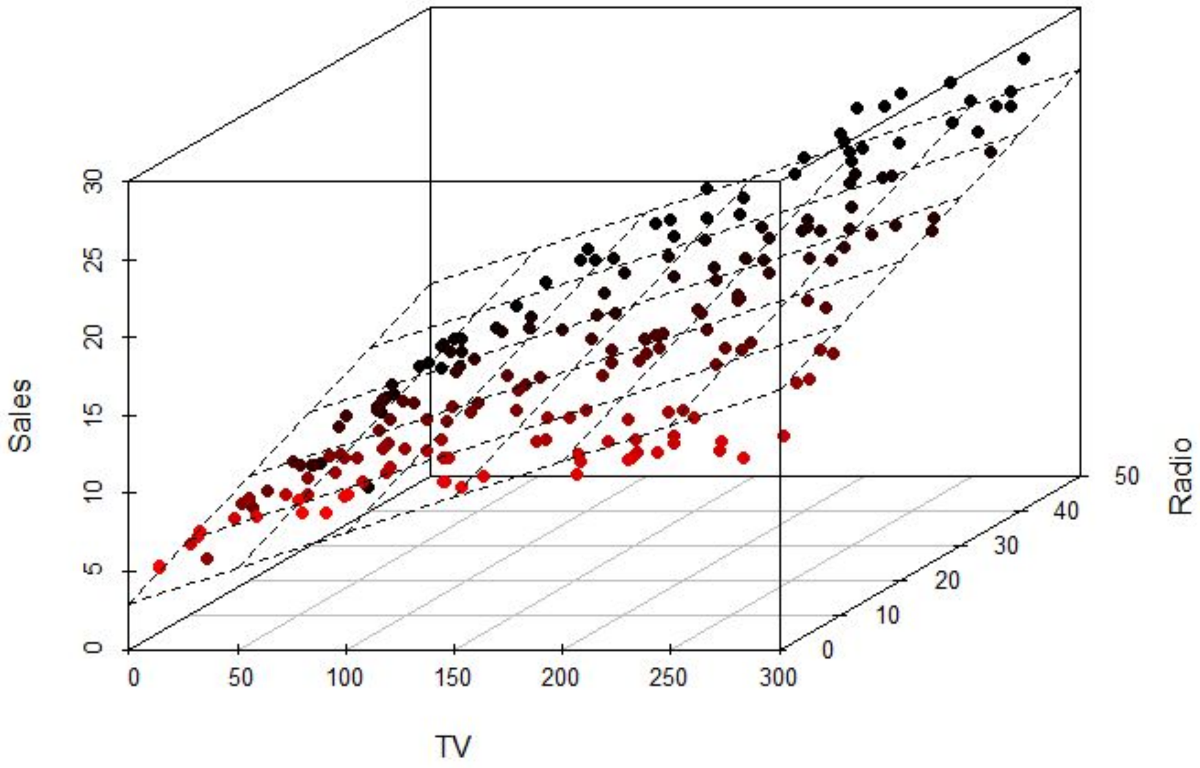
Regresión lineal: una variable

$$\underbrace{\boxed{\hat{y}}}_{\text{Valor predecido}} = \underbrace{\beta_0 + \beta_1}_{\text{Coeficientes}} \underbrace{\boxed{x}}_{\text{Entradas}}$$

Regresión lineal: múltiples variables

$$\boxed{\hat{y}} = \underbrace{\beta_0 + \beta_1}\boxed{x_1} + \ldots + \underbrace{\beta_p}\boxed{x_p}$$

boost
accelerate your learning



boost
accelerate your learning

Funciones de coste (Cost functions)

$$L_{OLS}(\hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i' \hat{\beta})^2 = ||y - X\hat{\beta}||^2$$

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i' \hat{\beta})^2 + \lambda \sum_{j=1}^{m} \hat{\beta}_j^2 = ||y - X\hat{\beta}||^2 + \lambda ||\hat{\beta}||^2.$$

$$L_{lasso}(\hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i' \hat{\beta})^2 + \lambda \sum_{j=1}^{m} |\hat{\beta}_j|.$$

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i=1}^{n} (y_i - x_i' \hat{\beta})^2}{2n} + \lambda \left(\frac{1 - \alpha}{2} \sum_{j=1}^{m} \hat{\beta}_j^2 + \alpha \sum_{j=1}^{m} |\hat{\beta}_j| \right),$$

Funciones de coste (Cost functions)

$$L_{OLS}(\hat{\beta}) = \sum_{i-1}^{n}(y_i - x_i'\hat{\beta})^2 = \left\|y - X\hat{\beta}\right\|^2$$

$$L_{ridge}(\hat{\beta}) = \sum_{i-1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda\sum_{j-1}^{m}\hat{\beta}_j^2 = \|y - X\hat{\beta}\|^2 + \lambda\left\|\hat{\beta}\right\|^2$$

$$L_{lasso}(\hat{\beta}) = \sum_{i-1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda\sum_{j-1}^{m}\left|\hat{\beta}_j\right|$$

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i-1}^{n}(y_i - x_i'\hat{\beta})^2}{2n} + \lambda(\frac{1-\alpha}{2}\sum_{j-1}^{m}\hat{\beta}_j^2 + \alpha\sum_{j-1}^{m}\left|\hat{\beta}_j\right|)$$

Métricas de evaluación para los modelos de regresión

Error cuadrático medio
(Mean Squared Error)

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

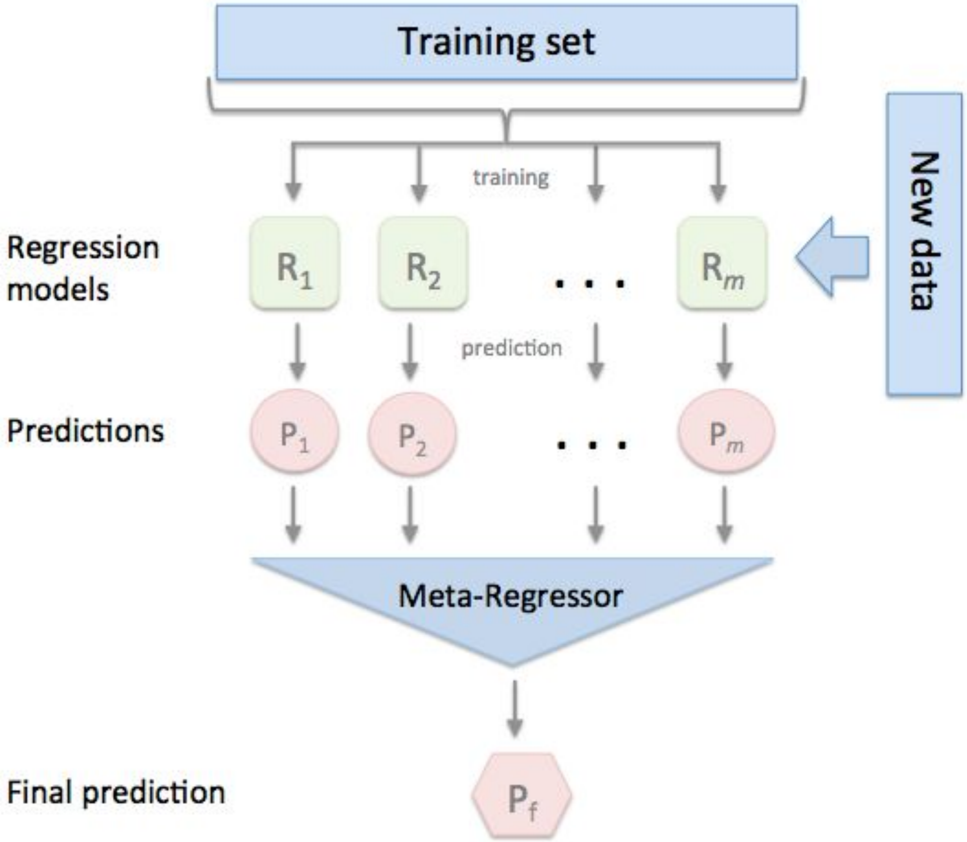
Raíz del error cuadrático medio
(Root Mean Squared Error)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$$

Media absoluta del error
(Mean Absolute Error)

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} |e_t|$$

Stacking de modelos



boost

accelerate your learning