

# Introducción a la Inteligencia Artificial

## Clase 4



## Índice

1. Notas clase anterior
2. Análisis de la regresión lineal ( $R^2$ )
3. Descomposición Bias-Variance
4. Práctica

## Coeficiente de determinación - “R cuadrado”

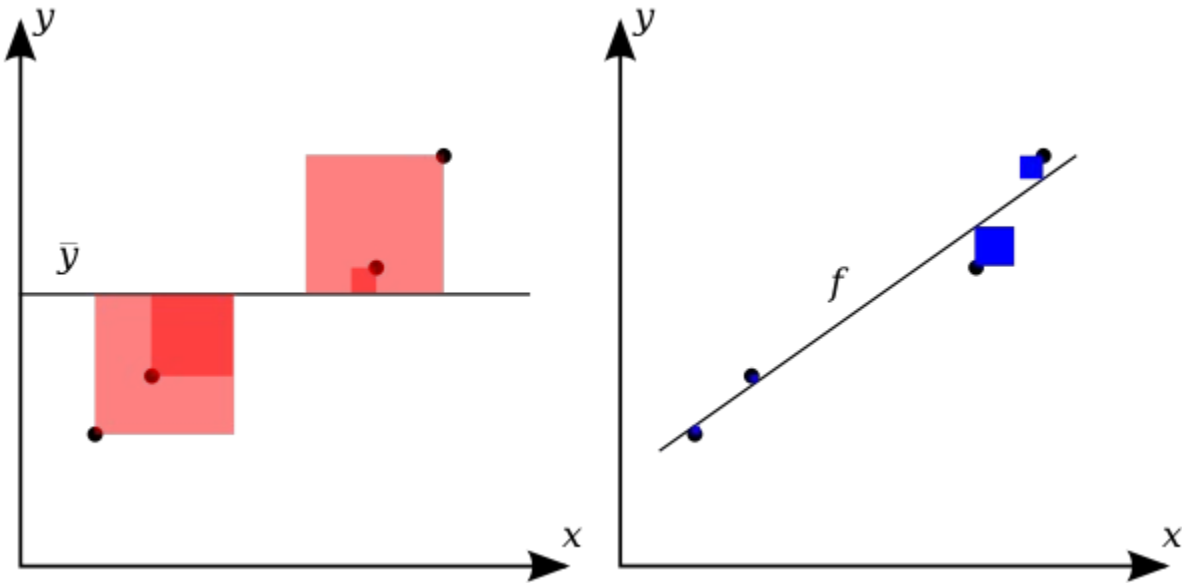
$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

SS = Sum of squares

res = residuos

tot = total

## Regresión Lineal - R2



$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

$$SS_{reg} = \sum_i (f_i - \bar{y})^2$$

$$SS_{res} = \sum_i (y_i - f_i)^2 = \sum_i e_i^2$$

$$\begin{aligned} SS_{tot} &= \sum_i (y_i - \bar{y})^2 \\ &= SS_{res} + SS_{reg} \end{aligned}$$

¿Similar a  $\sigma^2$ ?

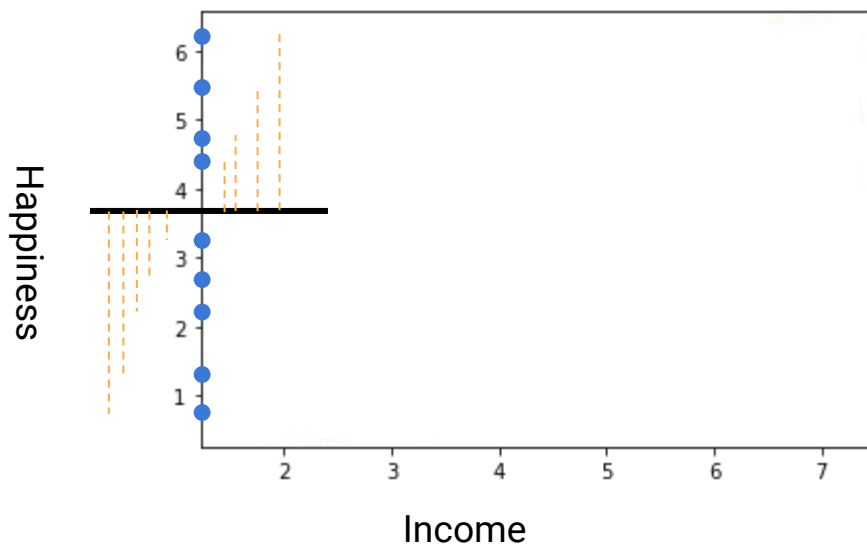
## Regresión Lineal - R2

$$\begin{aligned} R^2 &= 1 - \frac{SS_{res}}{SS_{tot}} \\ &= 1 - \left( \frac{SS_{res}}{SS_{tot}} * \frac{n}{n} \right) \\ &= 1 - \frac{\sigma_{res}}{\sigma_{tot}} \end{aligned}$$

Proporción de varianza no explicada

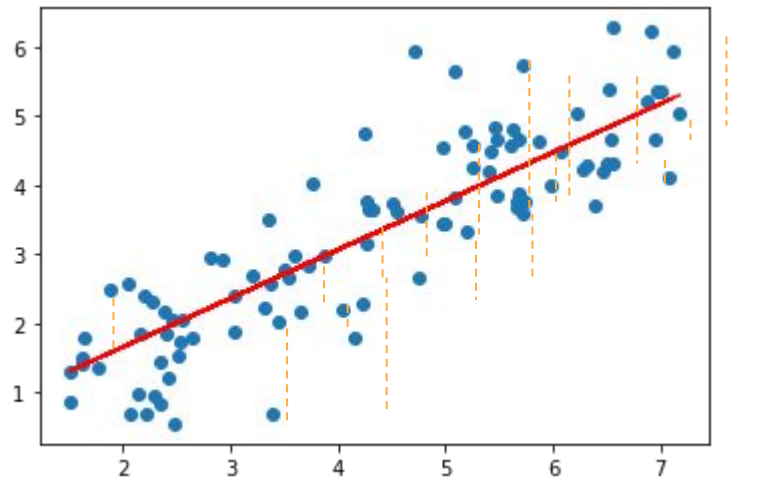
Proporción de varianza explicada

## Regresión Lineal - R2



$$SS(media) = (happiness - media)^2$$

$$Variación(media) = \frac{(happiness - media)^2}{n}$$

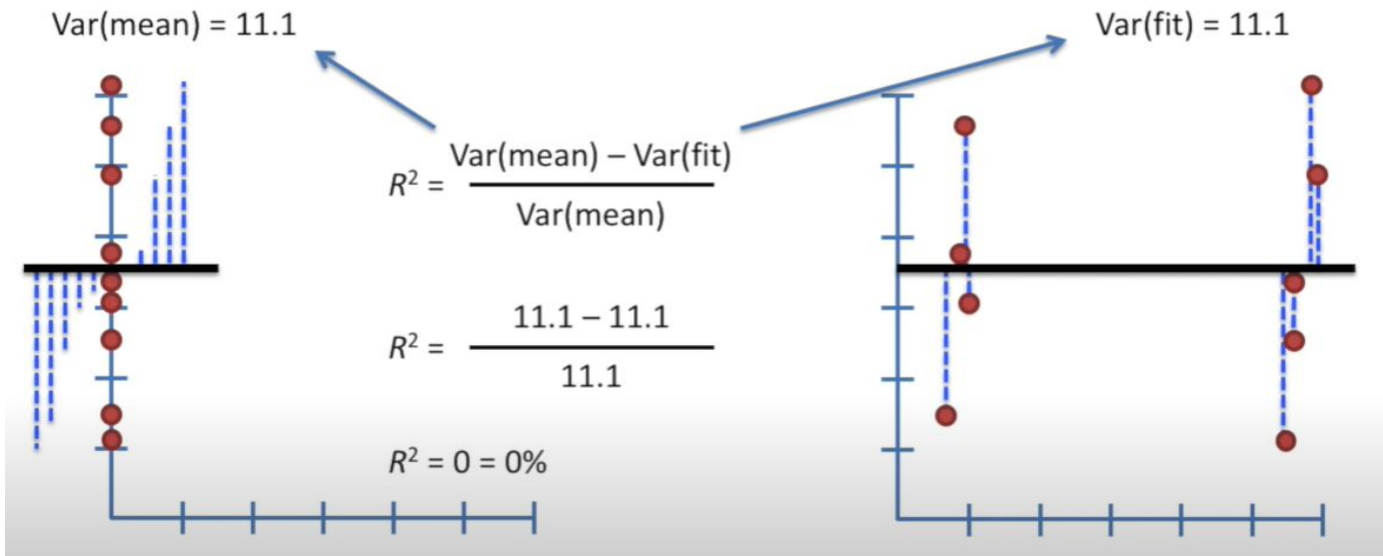


$$SS(fit) = (happiness - lr\_fit)^2$$

$$Variación(fit) = \frac{(happiness - lr\_fit)^2}{n}$$

Regresión Lineal - R2

$$R^2 = \frac{\text{Variación}(\text{media}) - \text{Variación}(\text{fit})}{\text{Variación}(\text{media})}$$



Fuente: StatQuest with Josh Starmer

## Regresión Lineal - R2

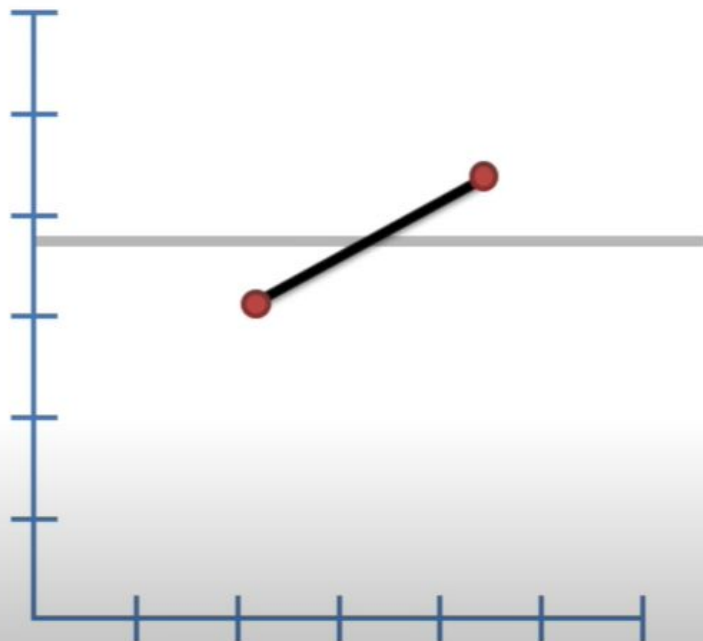
$$F = \frac{\text{Varación en happiness explicada por income}}{\text{Variación en happiness no explicada por income}}$$

$$SS(\text{mean}) = 10$$

$$SS(\text{fit}) = 0$$

$$R^2 = \frac{SS(\text{mean}) - SS(\text{fit})}{SS(\text{mean})}$$

$$= \frac{100 - 0}{100} = 100\%$$



Fuente: StatQuest with Josh Starmer



## R<sup>2</sup> y el coeficiente de correlación de Pearson

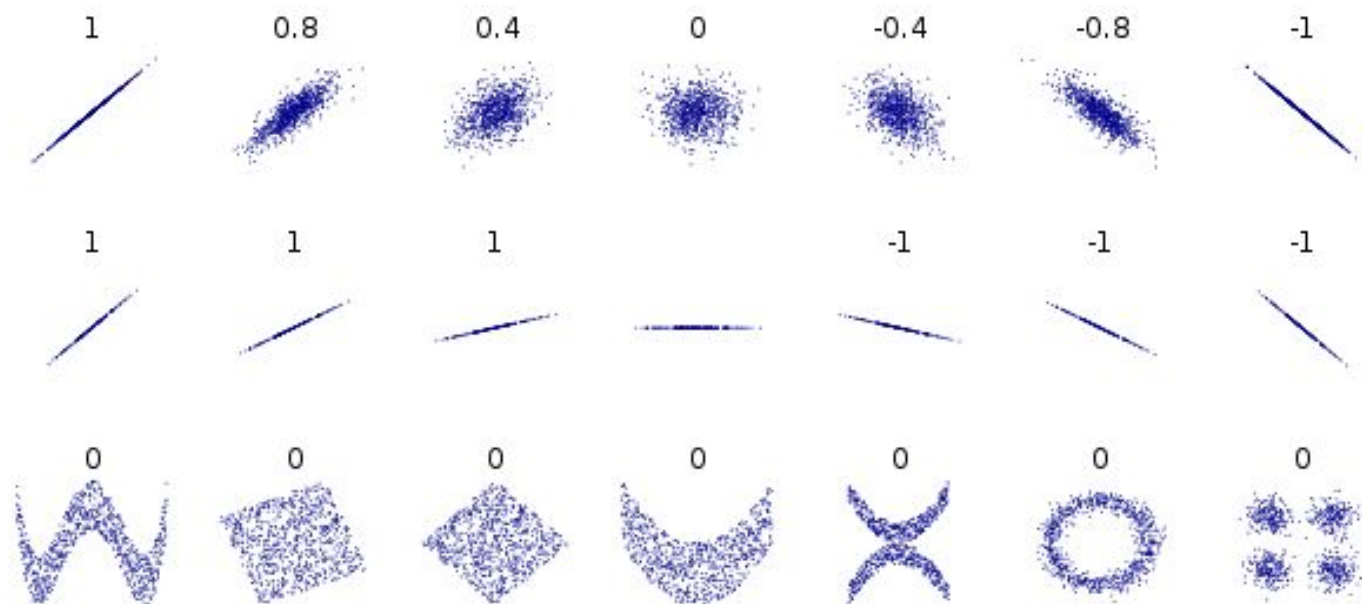
$$\rho = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \in [-1, 1]$$

Regresión múltiple **con ordenada** → Correlación entre observación y predicción

Regresión **con ordenada** → Correlación entre variable dependiente e independiente

$$\rho^2 = R^2 \in [0, 1]$$

## Coeficiente de correlación de Pearson

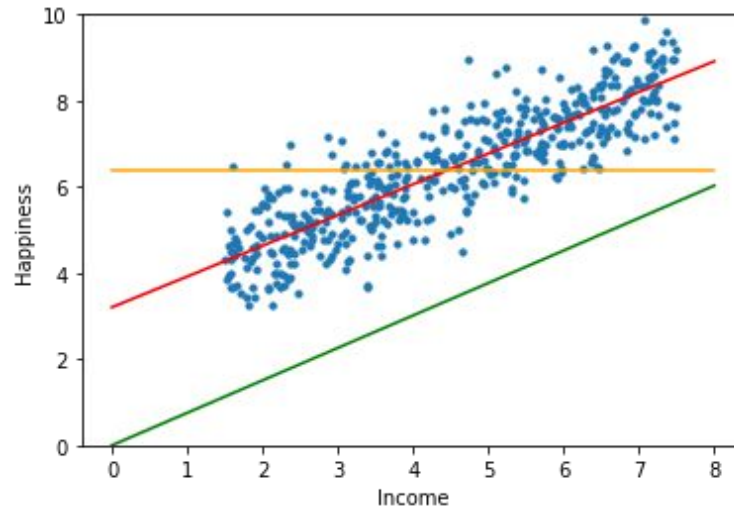


## R2 inflation

$\uparrow$  cantidad de predictores  $\rightarrow \uparrow R^2 \rightarrow$  F-test para comparación válida entre modelos.

## ¿R<sup>2</sup> negativo?

$$R^2 = 1 - \frac{\sigma_{res}}{\sigma_{tot}} < 0 \Leftrightarrow \sigma_{res} > \sigma_{tot}$$



Predecir con el promedio es mejor que el modelo

## Otras medidas a tener en cuenta

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}.$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

$$\text{RMSD} = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}}.$$

## Bias-Variance Tradeoff

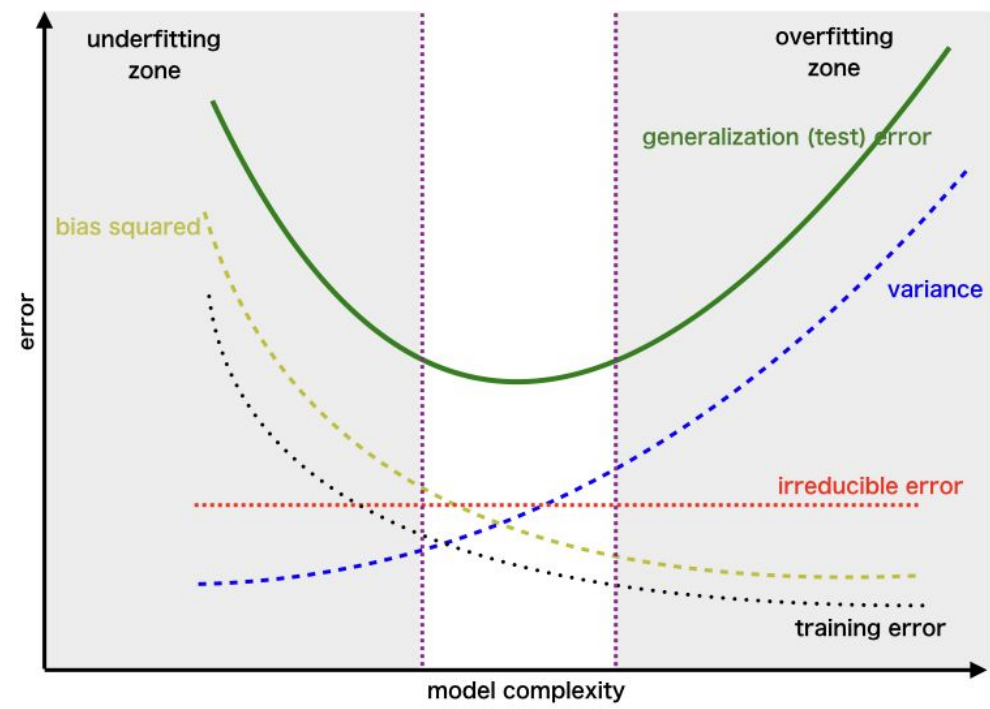
Cuando utilizamos el **error cuadrático medio** en un modelo de ML, podemos descomponer el mismo en términos de bias (sesgo) y variance (varianza).

$$MSE = Bias(\hat{f})^2 + Var(\hat{f}) + \sigma_\epsilon^2$$

$$Bias = E[\hat{f} - f]$$

$$Var(\hat{f}) = E[(E[\hat{f}] - \hat{f})^2]$$

## Bias-Variance Tradeoff



## Bibliografía

- The Elements of Statistical Learning | Trevor Hastie | Springer
- An Introduction to Statistical Learning | Gareth James | Springer
- Deep Learning | Ian Goodfellow | <https://www.deeplearningbook.org/>
- Stanford | CS229T/STATS231: Statistical Learning Theory | <http://web.stanford.edu/class/cs229t/>
- Mathematics for Machine Learning | Deisenroth, Faisal, Ong
- Artificial Intelligence, A Modern Approach | Stuart J. Russell, Peter Norvig