Introducción a la Inteligencia Artificial

Clase 4

.UBA fiuba

FACULTAD DE INGENIERÍA

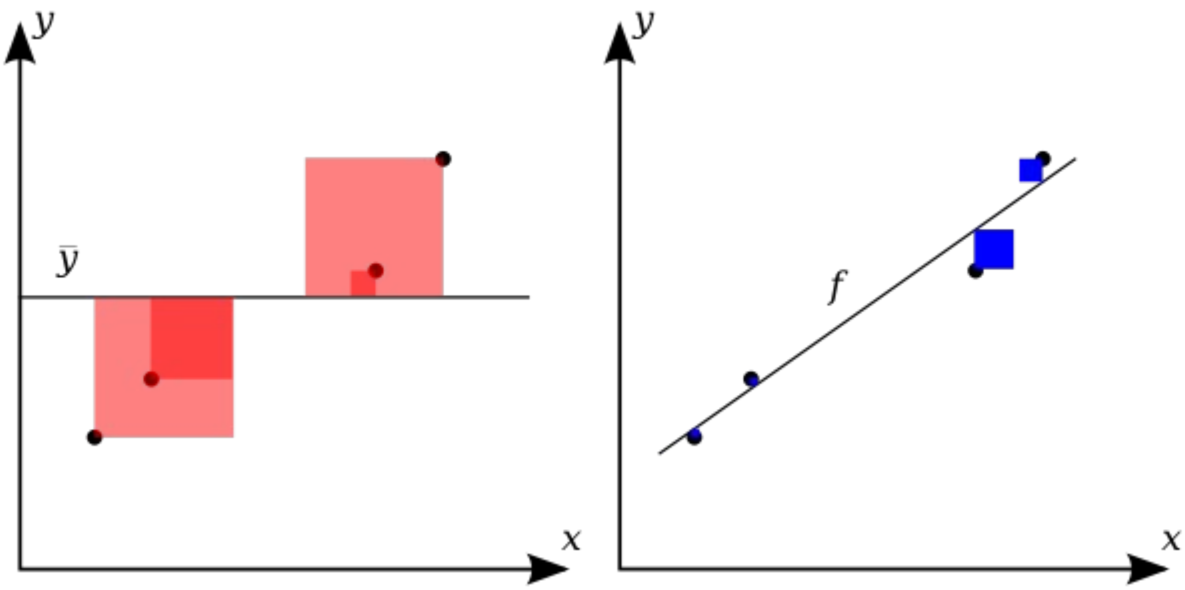
Índice

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



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

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

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

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

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

↳ i-esimo residuo

¿Similar a σ2?

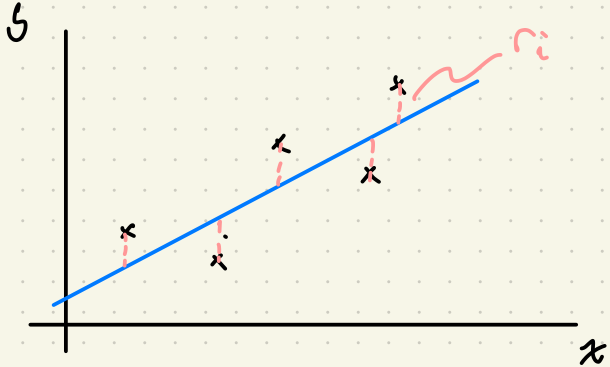
Regresión Lineal - R2

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

Proporción de varianza no explicada

Proporción de varianza explicada

• Analizamos inicialmente la regresión simple:



partimos de $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

definimos el i-esimo residuo como

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

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min \underbrace{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}_{f} \rightsquigarrow \min \sum_i r_i^2$$

$$\begin{cases} \partial_{\beta_0} f = 0 \to -2 \sum_i (y_i - \beta_0 - \beta_1 x_i) \longrightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \partial_{\beta_1} f = 0 \to -2 \sum_i (y_i - \beta_0 - \beta_1 x_i)\, x_i \longrightarrow \hat{\beta}_1 = \dfrac{\frac{1}{N} \sum (x_i - \bar{x})(y_i - \bar{y})^2}{\frac{1}{N} \sum_i (x_i - \bar{x})^2} \end{cases}$$

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{\hat{\mathrm{cov}}(x,y)}{\hat{\mathrm{var}}(x)} = \hat{\rho}_{xy} \end{cases}$$

→ Coeficiente de correlación de pearson

Vamos a analizar los errores:

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

↓ sumamos $-\bar{y}$ a ambos lados

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + r_i \rightsquigarrow \underbrace{\sum_{i=0}^{N} (y_i - \bar{y})^2}_{\text{TSS}} = \underbrace{\sum_{i=0}^{N} (\hat{y}_i - \bar{y})^2}_{\text{ESS}} + \underbrace{\sum_{i=0}^{n} r_i^2}_{\text{RSS}}$$

TSS Tasa de variabilidad total

ESS tasa de variabilidad explicada

RSS ↓ suma de residuos al cuadrado

Con TSS, ESS y RSS puedo construir: métricas, indices de bondad de ajuste o análisis de regresión avanzada

Con esto ¿Cómo sabemos si el modelo es bueno?

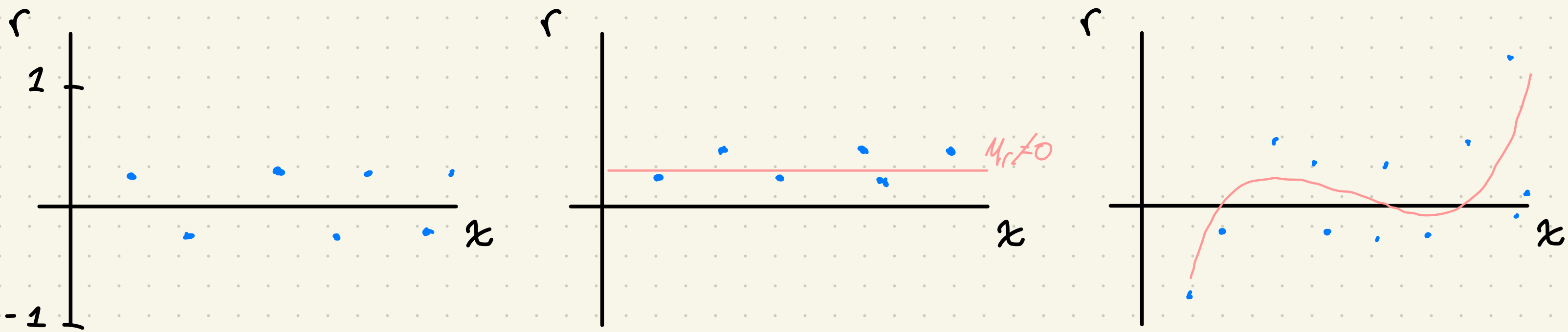
→ Error residual (RSE)

→ est. R^2 (coef. de pearson, R^2)

→ Estadístico F

$$+ \; RSE = \sqrt{\frac{RSS}{N-2}} = \frac{1}{N-2}\sqrt{\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}$$

si $RSE \downarrow \rightsquigarrow$ mi ajuste es bueno



RSE no es sensible a la distrib. de r

+ $R^2 \longrightarrow$ mide la bondad del ajuste
  ↳ equivalente al coef. de correlación al cuadrado:

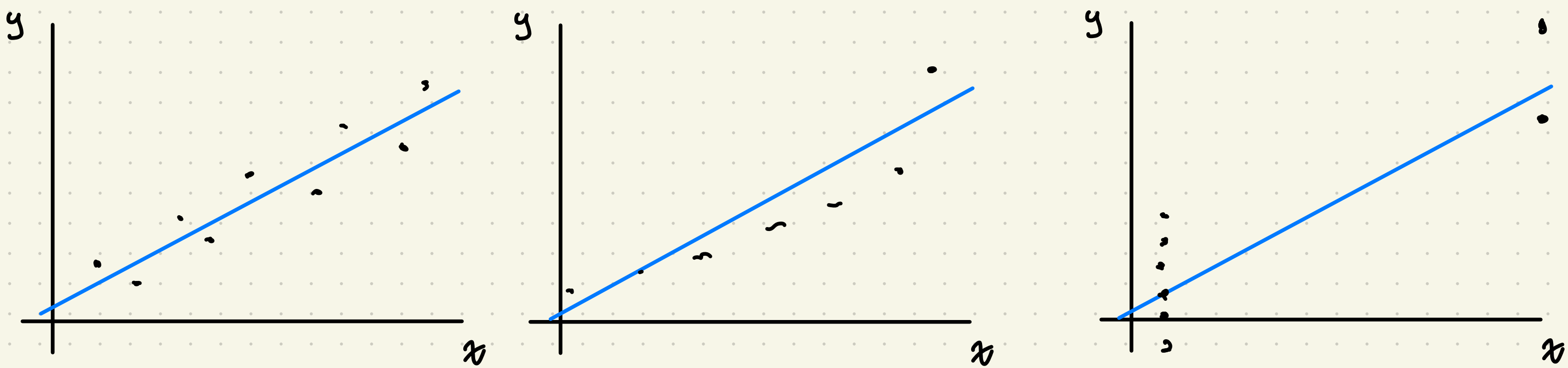
$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = \hat{\rho}_{XY}^{\,2}$$

(esto en regresión lineal simple)

$\rightsquigarrow$ R^2 no depende de las escalas (unidades) de y, solo de las proporciones

$\rightsquigarrow$ $R^2 \in [0,1] \rightarrow R^2 \sim 1$ $\rightsquigarrow$ el modelo es """"bueno""""

$R^2 \leq 0$ $\rightsquigarrow$ el modelo es peor que estimar con la media.



$\beta_0 \sim 3,\ \beta_1 \sim 0{,}5 \rightsquigarrow R^2 = 0{,}7$

- Estadístico F:

Voy a armar una tabla ANOVA del ajuste:

Fuentes de variación	SS	gl	cuad. medios
Explicada (ESS)	$\sum_i (\hat{y}_i - \bar{y})^2$	1	$\sum_i (\hat{y}_i - \bar{y})^2$
Residual (RSS)	$\sum r_i^2$	$n-2$	$S_R = \frac{1}{n-2} \sum_i r_i^2$
total (TSS)	$\sum_i (y_i - \bar{y})^2$	$n-1$	

construimos $F = \dfrac{ESS}{S_R^2}$ $\rightarrow$ si $F \uparrow$ $\rightsquigarrow$ la variabilidad explicada es muy grande respecto a la no explicada

Planteamos el sig. test de hipótesis:

$$TH: \begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases} \rightsquigarrow \mathbb{P}(F \geq f_{(gl,\alpha)})$$

$$f_{1, n-2, 1-\alpha}$$

¿Cómo interpretamos esto?

- hipotesis nula: supone que mi OLS no es mejor que un modelo constante. no explica mejor la varianza que un modelo cte.
- hipótesis alternativa: mi modelo OLS explica mejor (es estadísticamente significativo) la varianza de la respuesta que un modelo constante.

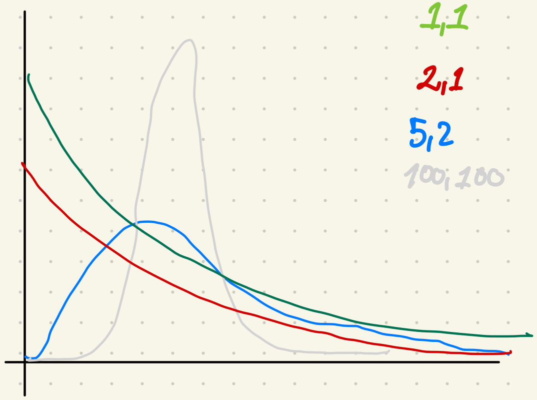
¿Cómo calculamos el F-test?

1. Armamos la tabla anova.

2. obtenemos el estadístico $F = ESS/S_r^2$

3. calculamos el p-valor, f_{crit} ⇝ definimos el nivel de significancia α.

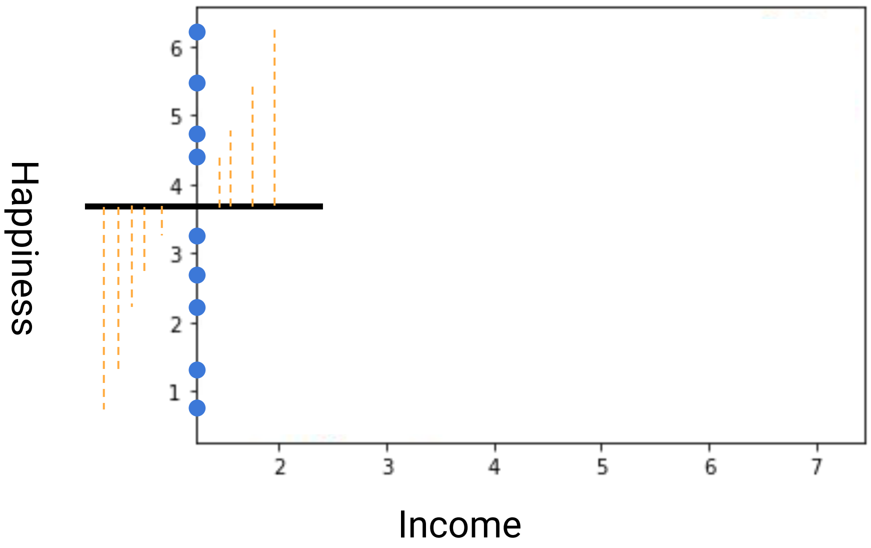
$$f \sim \mathcal{F}_{gl_{ESS},\, gl_{RSS}}$$

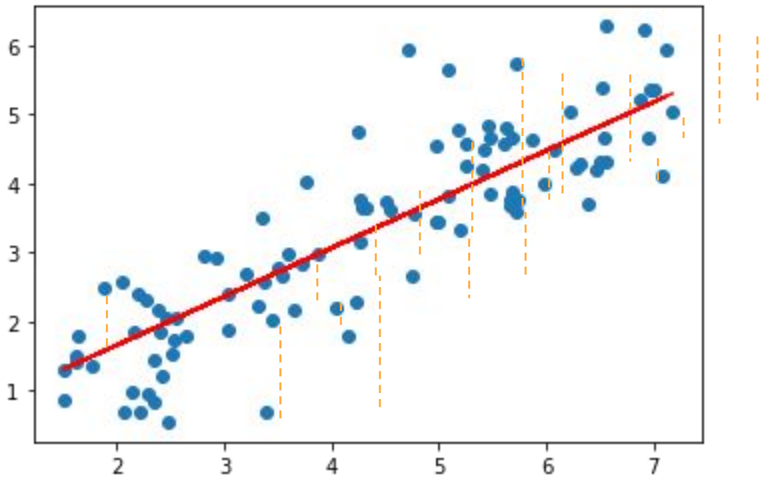


4. buscar $\mathbb{P}(\mathcal{F} \geq f_{crit}) = \alpha$

5. si $F > f_{crit} \Rightarrow$ rechazo H_0

Regresión Lineal - R2





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

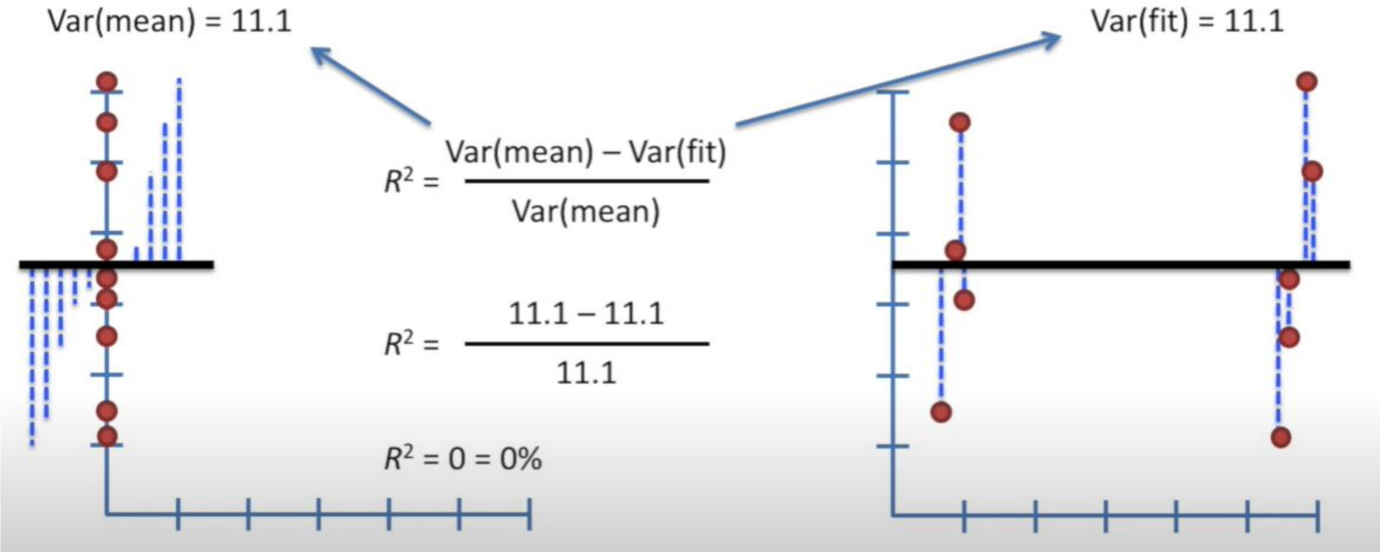
$$Variación(media) = \frac{(happiness - media)^2}{n}$$

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

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

Regresión Lineal - R2

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



Fuente: StatQuest with Josh Starmer

Regresión Lineal - R2

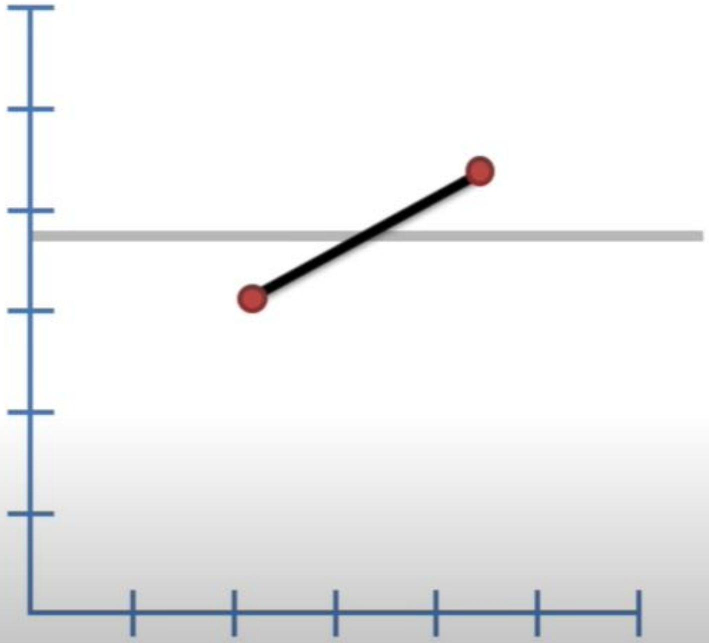
$$F = \frac{Varación\ en\ happiness\ explicada\ por\ income}{Variación\ en\ happiness\ no\ explicada\ por\ income}$$

SS(mean) = 10

SS(fit) = 0

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

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



Fuente: StatQuest with Josh Starmer

R2 y el coeficiente de correlación de Pearson

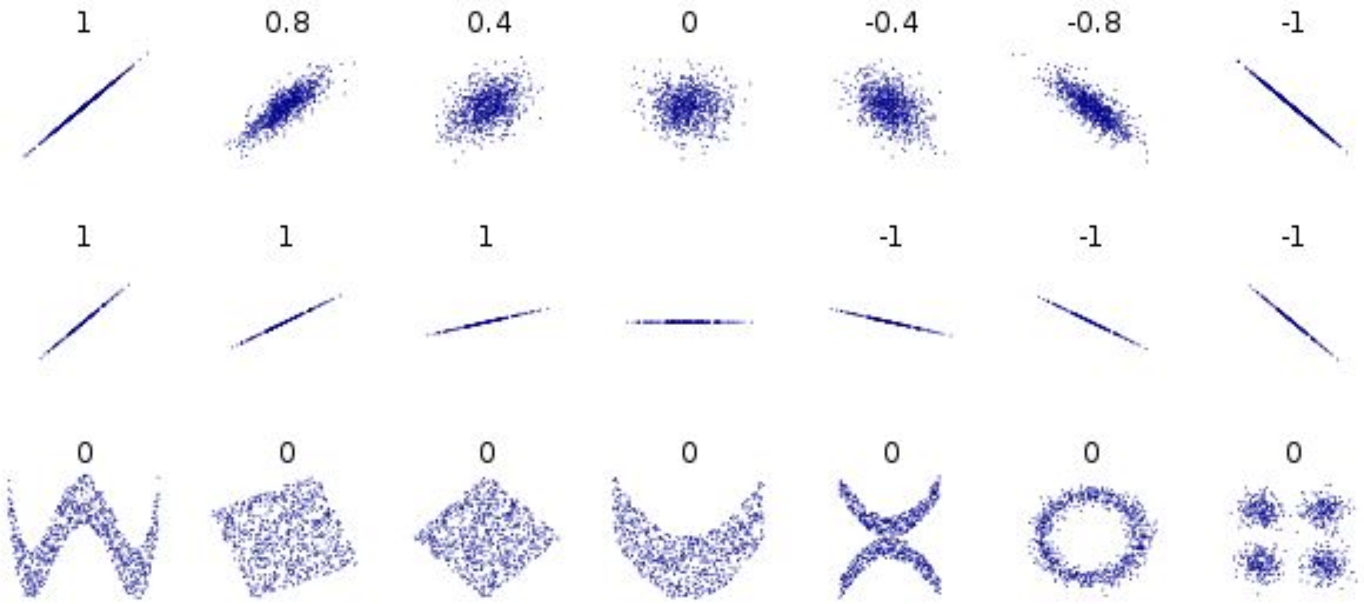
$$\rho = \frac{cov(x,y)}{\sigma_x \sigma_y} \in [-1,1]$$

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

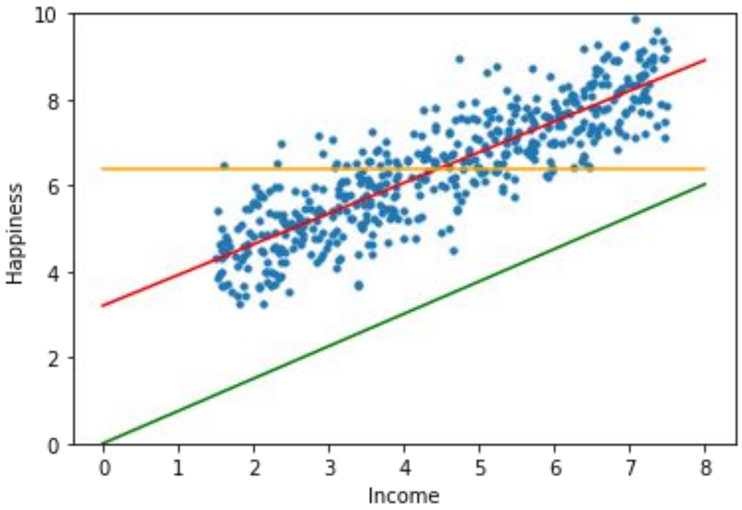


R2 inflation

$\uparrow$ cantidad de predictores $\longrightarrow \uparrow$R2$\longrightarrow$ F-test para comparación válida entre modelos.

¿R2 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

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

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

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

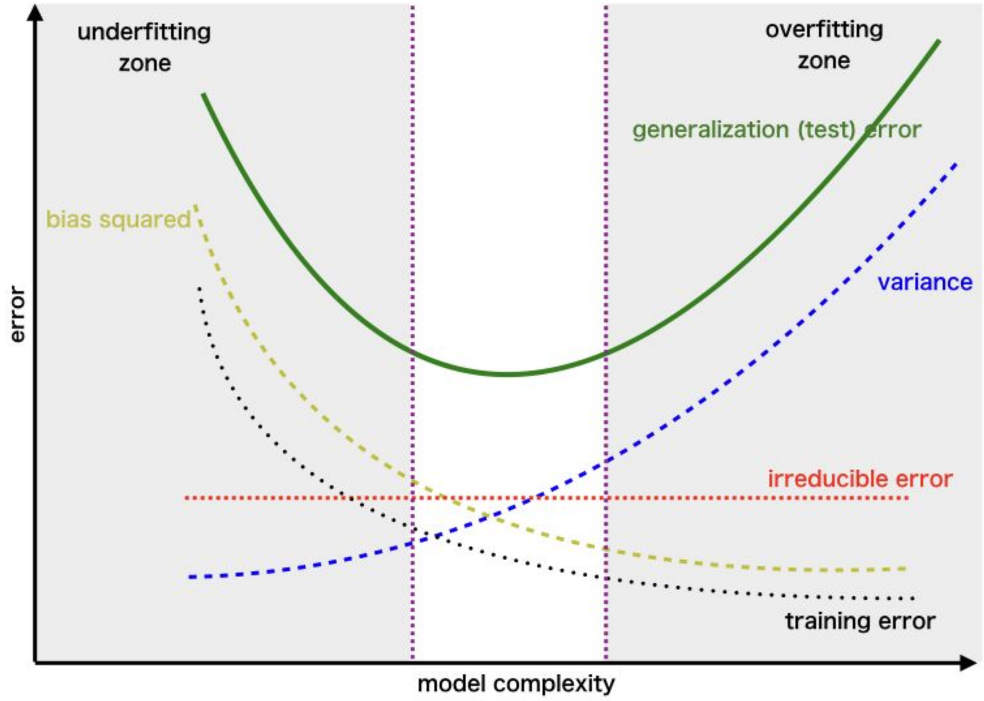
Bias-Variance Tradeoff

Cuando utilizamos el **error cuadrático medio** en un modelo de ML, podemos descomponer el mísmo 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