

Modelos Estadísticos para Clasificación y Regresión

Práctico 3 - Sesgo y Varianza

IMERL - FIng

6 de setiembre de 2023



1 Repaso

2 Práctico 3

1 Repaso

Error verdadero vs. error en *test*

Objetivo

Sesgo y varianza

2 Práctico 3

1 Repaso

Error verdadero vs. error en *test*

Objetivo

Sesgo y varianza

2 Práctico 3



Error verdadero vs. error en *test*

Error verdadero

Sea D la distribución de los pares (x, y) , y sea \hat{f} el modelo entrenado con el *dataset* S . Definimos el error verdadero como

$$MSE_v = E(y - \hat{f}(x))^2$$

Error en *test*

Sea \hat{f} el modelo entrenado con el *dataset* S y sean (x_{test}, y_{test}) los datos de *test*. Definimos el error en *test* como

$$MSE_t = E(y_{test} - \hat{f}(x_{test}))^2$$



1 Repaso

Error verdadero vs. error en *test*

Objetivo

Sesgo y varianza

2 Práctico 3

Objetivo

Queremos reducir el error en *test*:

$$E \left(y_0 - \hat{f}(x_0) \right)^2$$

donde

- x_0, y_0 es un dato de *test*.
- \hat{f} es el modelo entrenado con el *dataset* S .



1 Repaso

Error verdadero vs. error en *test*

Objetivo

Sesgo y varianza

2 Práctico 3



Descomposición *sesgo-varianza*

El error en *test* se puede descomponer en:

$$E \left(y_0 - \hat{f}(x_0) \right)^2 = \text{Var}(\hat{f}(x_0)) + [\text{Bias}(\hat{f}(x_0))]^2 + \text{Var}(\epsilon).$$

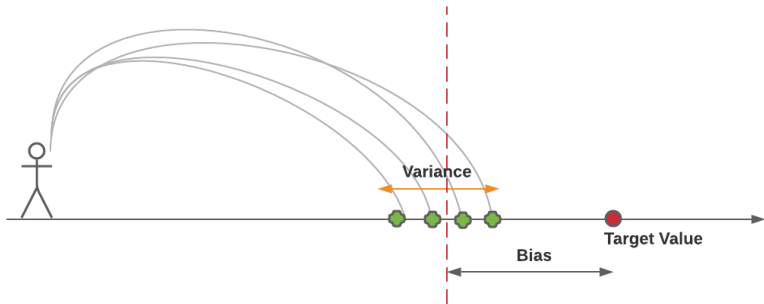
donde

- $\text{Var}(\hat{f}(x_0))$ es la *varianza* del algoritmo de aprendizaje, y representa la sensibilidad del error a pequeños cambios en S (lo ideal sería que \hat{f} varíe poco con diferentes *datasets*).
- $\text{Bias}(\hat{f}(x_0))$ es el *sesgo* del alg. de aprendizaje, y proviene de hipótesis erróneas y aproximaciones introducidas al elegir el modelo.
- $\text{Var}(\epsilon)$ es el *error irreducible*, y proviene del ruido que contienen las etiquetas y_i (independiente de nuestro algoritmo).

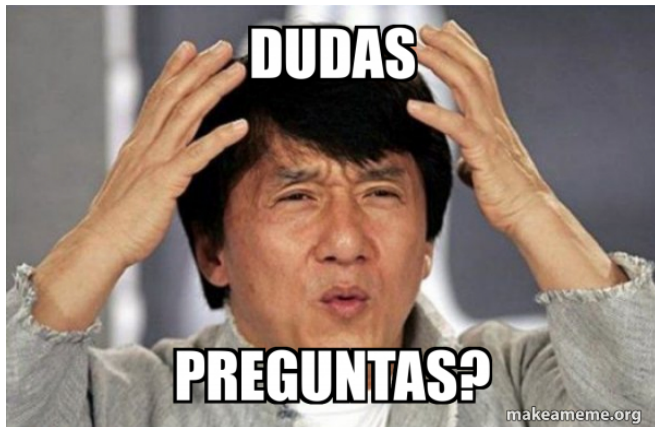
Compromiso *sesgo-varianza*

Como el error irreducible es irreducible, interesa minimizar $Var(\hat{f}(x_0))$ y $Bias(\hat{f}(x_0))$.

Se habla de que existe un *compromiso* entre sesgo y varianza porque es relativamente fácil obtener un método con bajo sesgo y alta varianza (ej: polinomio interpolante) o un método con baja varianza pero alto sesgo (ej: recta horizontal).



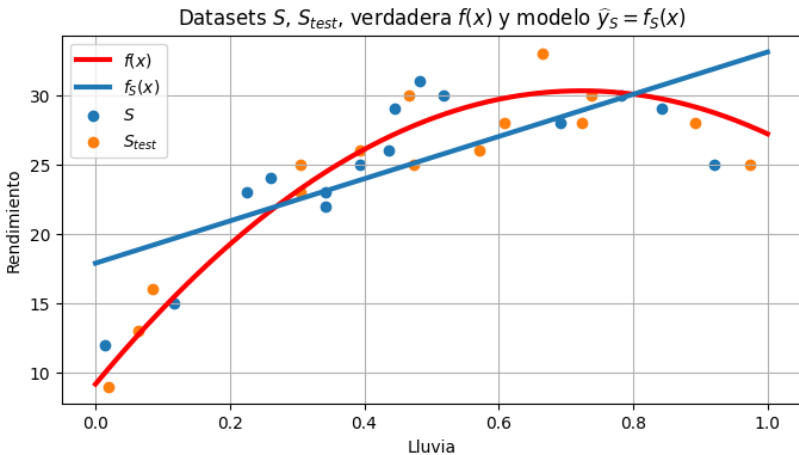
Preguntas?



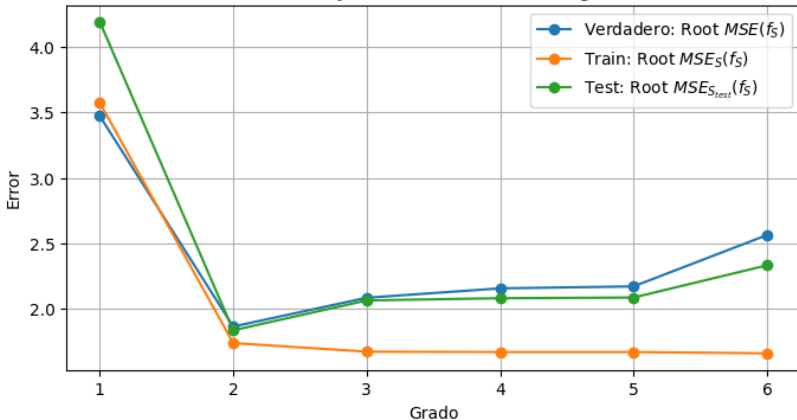
1 Repaso

2 Práctico 3

Práctico 3



Práctico 3

Errores de $\hat{y}_S = f_S(x)$ en función del grado

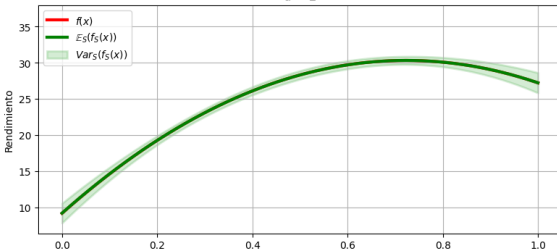
Práctico 3

Sesgo y Varianza

d = 1



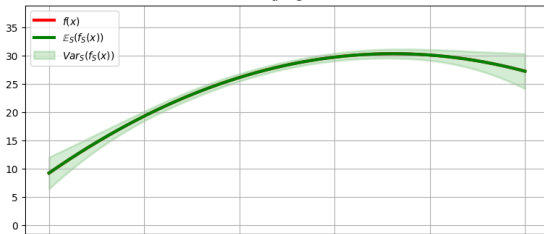
d = 2



Práctico 3

Sesgo y Varianza

d = 3



d = 4

