


Hiperparámetros y parámetros

Hiperparámetros: Características de un algoritmo de ML que definen la forma como se lleva a cabo el entrenamiento.

Ej: Hiperparámetros de KNN : - Número de vecinos
- Peso para las distancias
:

Hiperparámetros de árboles de decisión : - Profundidad máx
- # máx de hojas
:

(Ojo En sklearn se llaman "Parámetros" no confundir)
↓
son los parámetros del algoritmo de ML

Parámetros: Describen completamente al modelo arrojado por el algoritmo de M. Son los que se determinan o ajustan en el entrenamiento

Ej : - Coeficientes e intercepto de la regresión lineal
- Pesos y Sesgo en la red neuronal

Varianza, Sesgo, Sobreajuste, Subajuste

Sea D un conjunto de datos $D = \{ (\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_n, y_n) \}$

Denotemos por A a un algoritmo de ML.

$$h_D = A(D)$$

Modelo obtenido luego de entrenar en D

Predicción
para x

$$h_D(x) = \hat{y}$$

Sesgo y Varianza

Valores asociados a A
con sus hiperparámetros
y a diferentes conjuntos de
entrenamiento

Sobreajuste y Subajuste

Concepto asociado a un algoritmo A
y a un conjunto de datos fijo
a lo largo del aprendizaje o
entrenamiento.

(no es un valor o número)

Sea f una función que describe el comportamiento de un fenómeno del cual tenemos datos

$$Y = f(X) + \varepsilon \quad E(\varepsilon) = 0 \quad \text{Var}(\varepsilon) = \sigma_\varepsilon^2$$

Err := Error esperado de predicción, dado un algoritmo A

$$\text{Err} := E_{\substack{x, y, D \\ \downarrow \text{Datos de Test} \quad \downarrow \text{Conjunto de entrenamiento}}} \left[(y - h_D(x))^2 \right]$$

Sea f una función que describe el comportamiento de un fenómeno del cual tenemos datos

$$Y = f(x) + \varepsilon \quad E(\varepsilon) = 0 \quad \text{Var}(\varepsilon) = \sigma_\varepsilon^2$$

Err := Error esperado de predicción, dado un algoritmo A

$$\text{Err} := E_{\substack{x, y, D \\ \text{Datos de Test} \quad \text{Conjunto de entrenamiento}}} \left[(y - h_D(x))^2 \right]$$

Err presenta la siguiente descomposición:

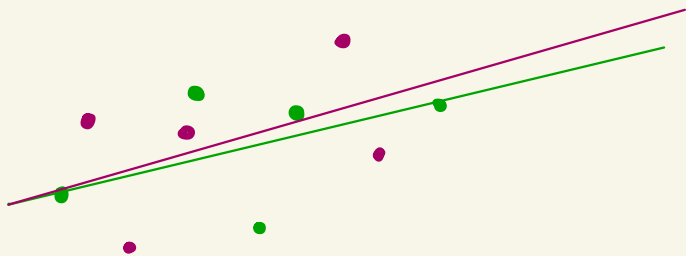
$$\text{Err} = \overbrace{E_{x, D} \left[(h_D(x) - \bar{h}(x))^2 \right]}^{\text{VARIANZA}} + \overbrace{E_x \left[(\bar{h}(x) - \bar{y}(x))^2 \right]}^{\text{SESGO}^2} + E_{x, y} \left[(\bar{y}(x) - y)^2 \right]$$

↓ Modelo dado por A para D
↓ Modelo esperado de A
↓ Promedio de predicciones del modelo
 ↓ Media de valores reales

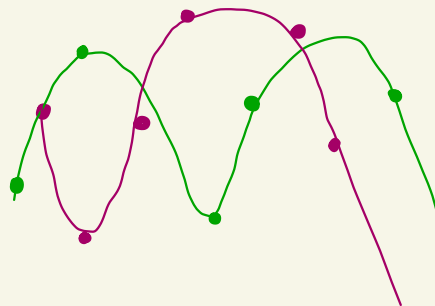
Varianza en palabras

$$E_{x,D}[(h_D(x) - \bar{h}(x))^2]$$

Mide qué tanto varía el modelo dado por el algoritmo A al cambiar los datos de entrenamiento



baja varianza

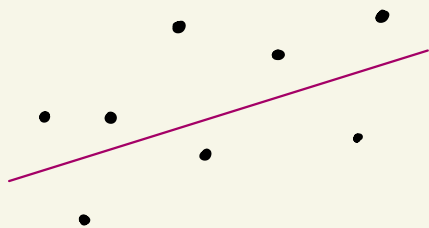


alta varianza

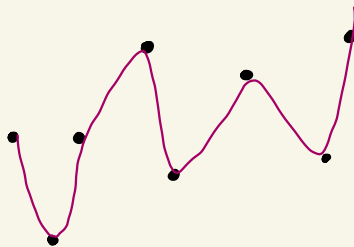
Sesgo en palabras

$$E_x [(\bar{h}(x) - \bar{y}(x))]^2$$

Que tan lejos está el modelo promedio, de los datos reales



Mayor Sesgo



Menor Sesgo

Complejidad del modelo

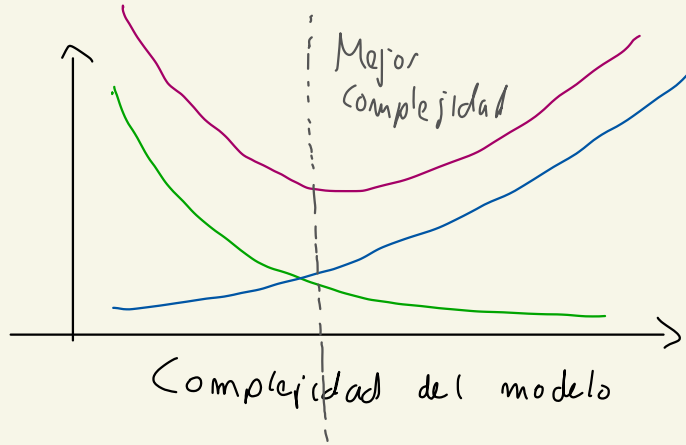
Arbol de decision: Profundidad del árbol, # de hojas

KNN # de vecinos

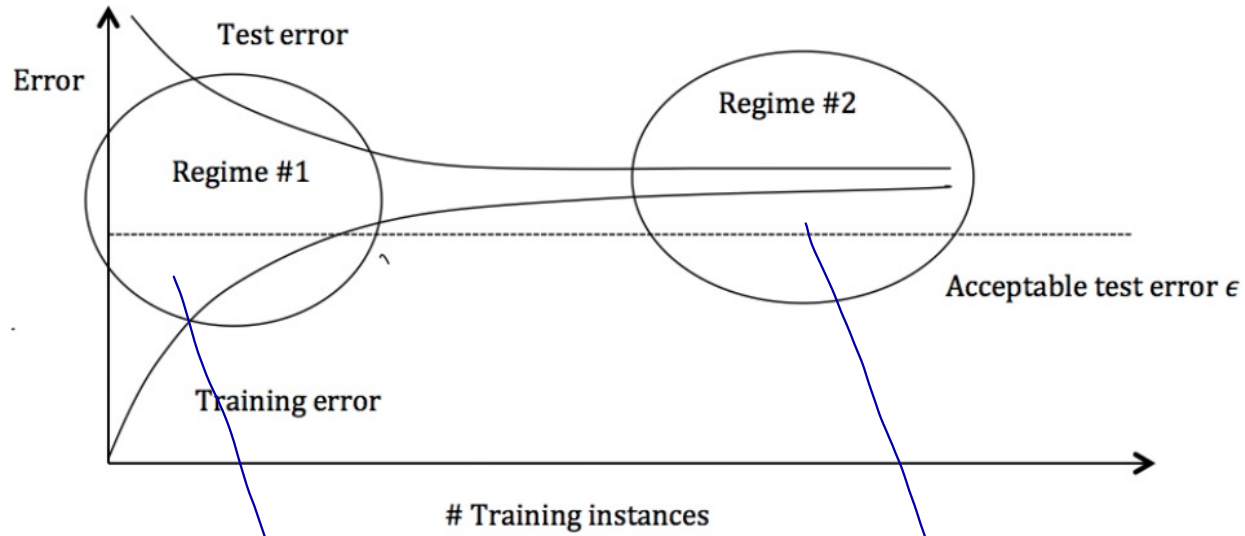
Redes neuronales # Capas, # unidades en las capas
"neuronas"
"nodos"

Linear regression: Tamaño del subconjunto de características

Entre más complejo sea el modelo, la varianza tiende a aumentar y el sesgo a disminuir



Err
Varianza
Sesgo



Alta varianza

Alto sesgo

?

?

Varianza

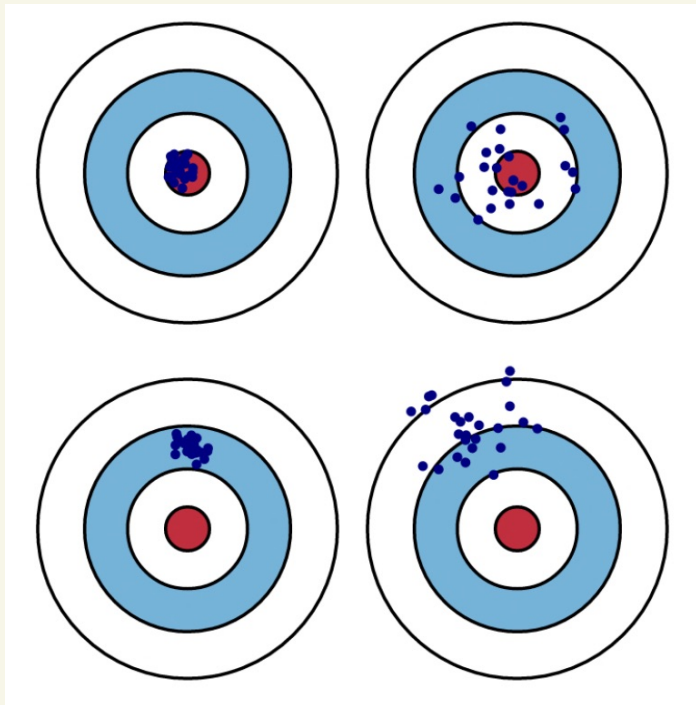
$$E_{x,D}[(h_D(x) - \bar{h}(x))^2]$$

Sesgo

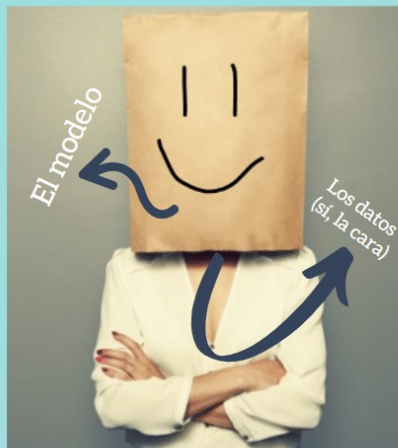
$$E_x[(\bar{h}(x) - \bar{y}(x))^2]$$

?

?



ALTO SESGO BAJA VARIANZA



El modelo tiene un error alto con respecto a los datos de entrenamiento

Sin embargo...

El error no varía mucho al cambiar los datos de entrenamiento.

No se ajusta perfectamente... se comportará de forma similar con cualquier otra persona.

El modelo se ajusta bastante bien a los datos de entrenamiento

Pero...

Difícilmente se ajusta a otros datos (como los datos de prueba)

¡Máscara perfecta!... pero probablemente sólo se ajustará a la Sra. Doubtfire

ALTA VARIANZA BAJO SESGO



...

Sobreajuste y subajuste

Tenemos un algoritmo A con sus hiperparámetros definidos

Sobreajuste

El modelo aprende en exceso las particularidades de los datos de entrenamiento y pierde la capacidad de generalizar

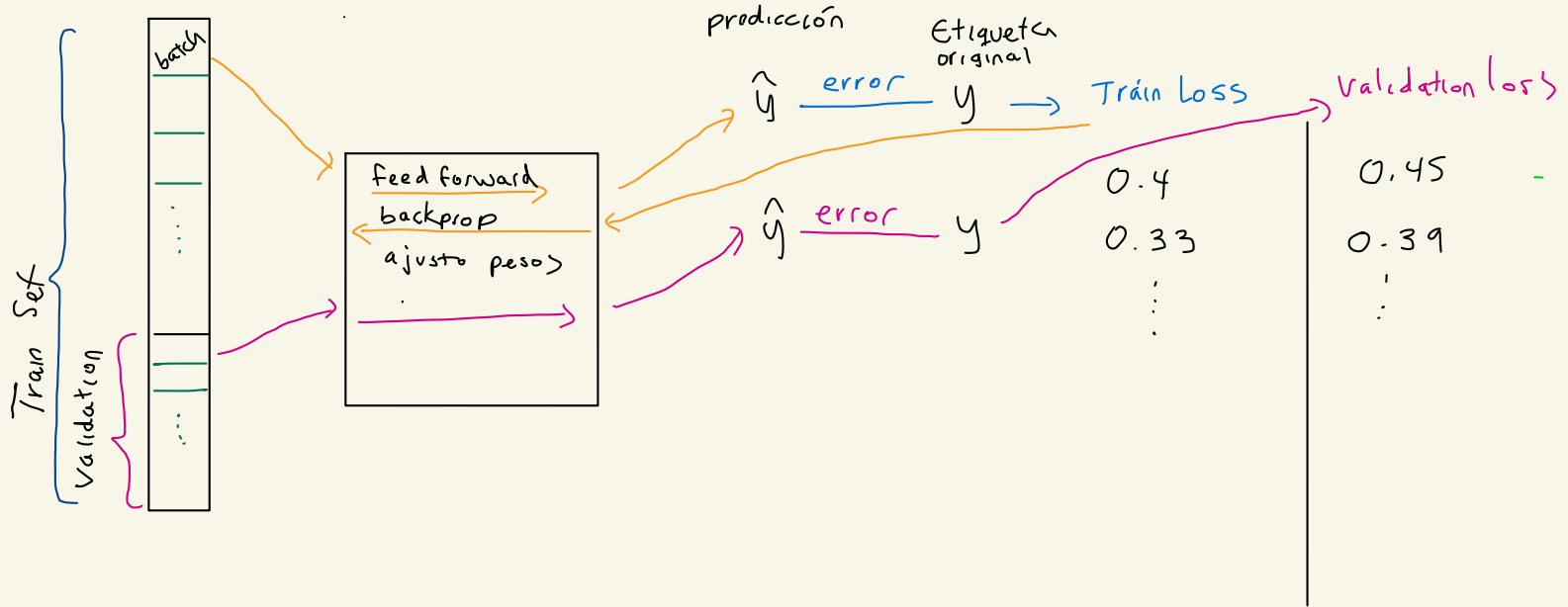
Un modelo con alta complejidad (alta varianza) tiende a sobreajustar

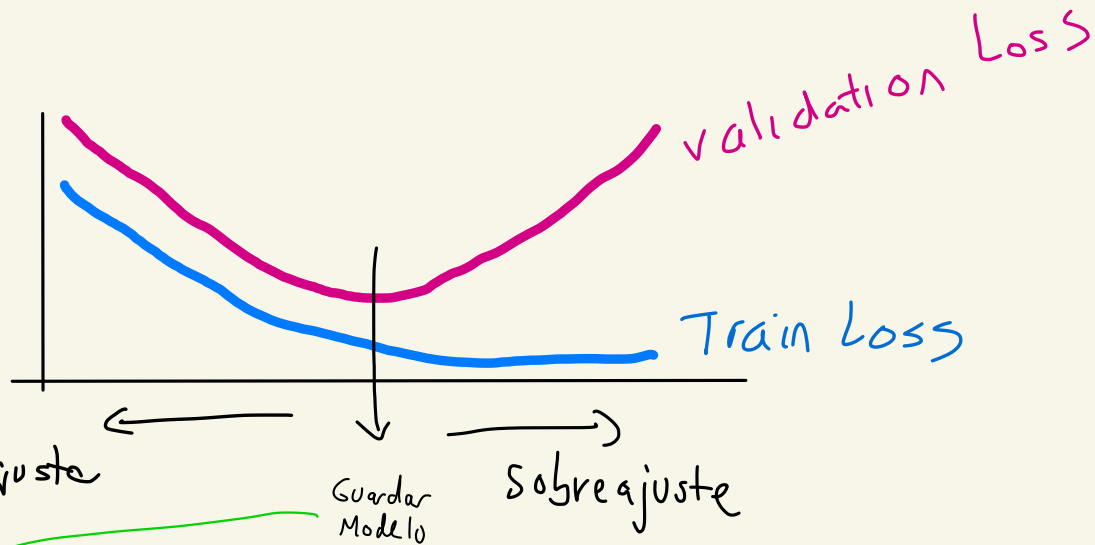
Subajuste

El modelo es muy simple para los datos, por lo tanto no logra predecir bien.


Un modelo con baja complejidad tiende a subajustar

Para medir el sobreajuste/subajuste durante el entrenamiento de un modelo, se utiliza un conjunto de validación





Test Set



← Prueba definitiva

Evitar el sobreajuste

- ☐ Regularización (Ridge, Lasso)
- ☐ Early stopping

Evitar el subajuste

- ☐ Incrementar la complejidad del modelo
- ☐ Más datos de entrenamiento, más características

Otros (más adelante)