

MSI603 – Proyecto Integrador: Ciencia de Datos



**Universidad
Andrés Bello®**
Conectar · Innovar · Liderar



Mag. en Ingeniería Informática

Entreno al modelo con
1 sólo raza de perro

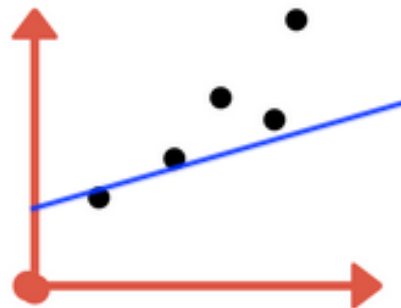


Muestra nueva:
¿Es perro?



NO

FALLO



underfitting

Entreno al modelo con
10 razas de perro color marrón

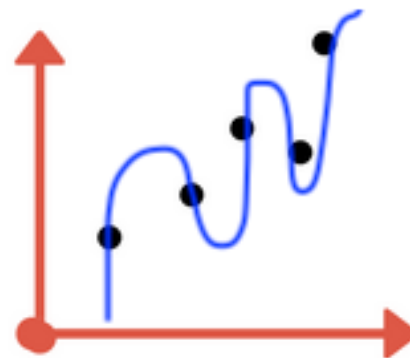


Muestra nueva:
¿Es perro?



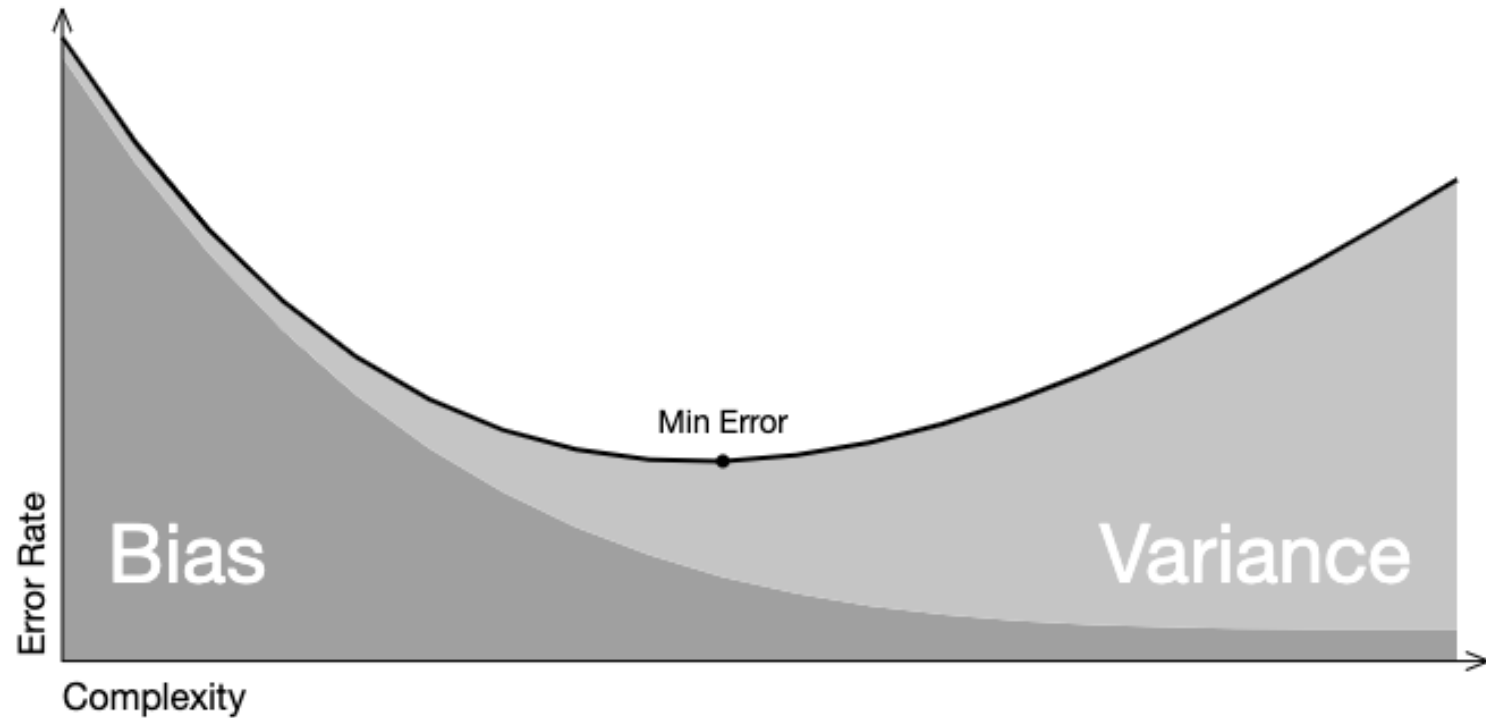
NO

FALLO



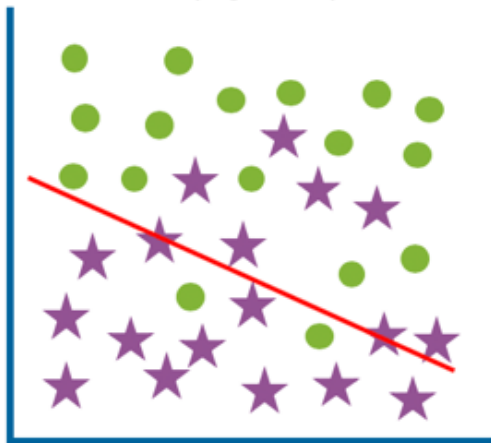
overfitting

Equilibrio entre sesgo y sobreajuste



Subajuste vs Sobreajuste

Underfit
(high bias)



High training error
High test error

Optimum



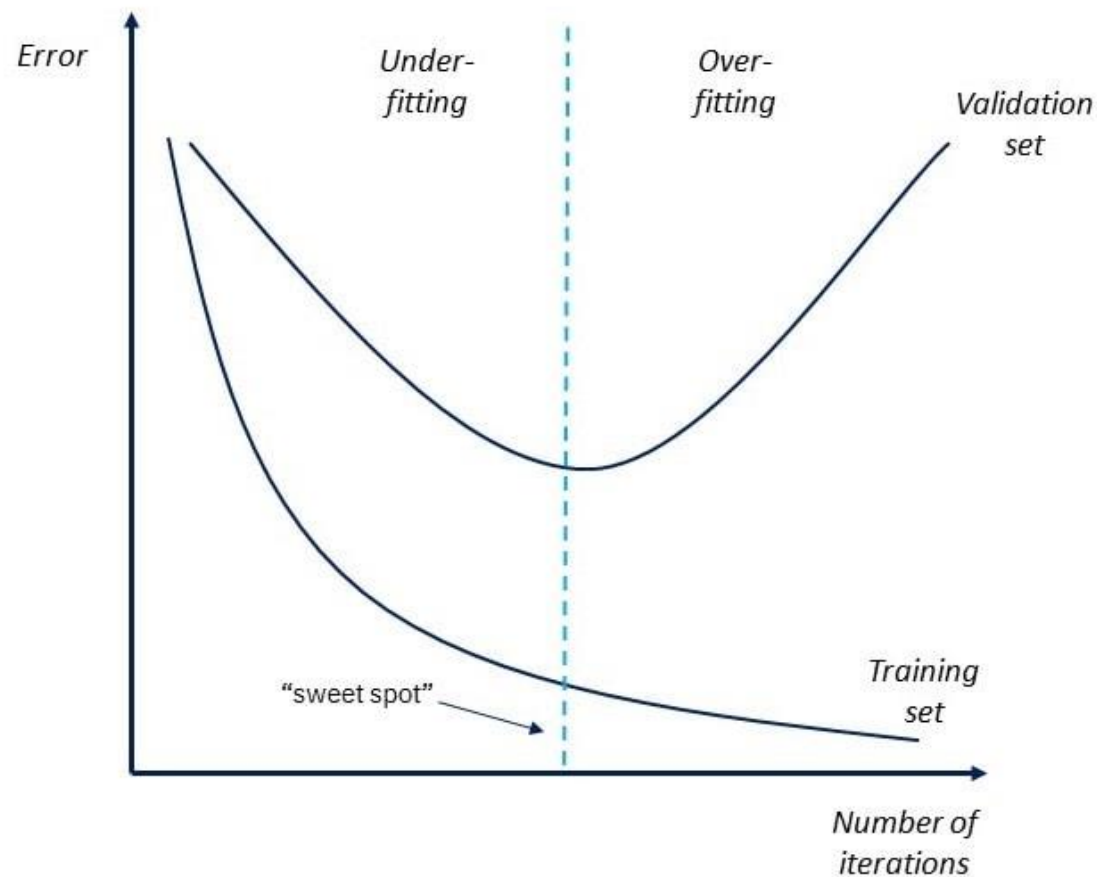
Low training error
Low test error

Overfit
(high variance)



Low training error
High test error

Detención temprana (Early Stopping):



- Imputación de datos para entrenar con un conjunto de datos más grande.
- Preferentemente es mejor aumentar el tamaño del conjunto de datos sin agregar datos sintéticos.
- Selección de columnas/características relevantes.

- Regularización L1: se aplica una penalización a coeficientes más grandes.
- Dropout: aleatoriamente anula un porcentaje de entradas desde la capa anterior a la siguiente.