

# Módulo 2: Conceptos generales de Machine Learning

## *2.4. Conceptos de Overfitting y Underfitting*

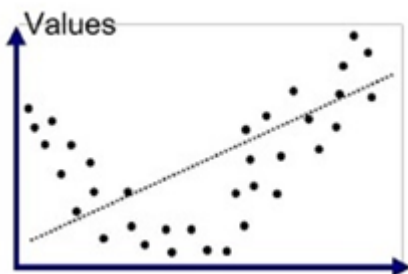
---

Rafael Zambrano

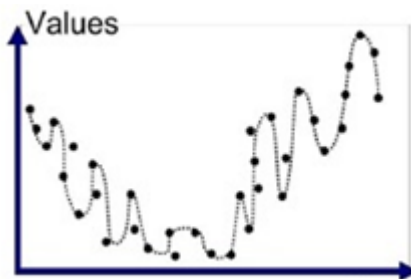
[rafazamb@gmail.com](mailto:rafazamb@gmail.com)

# Overfitting y Underfitting

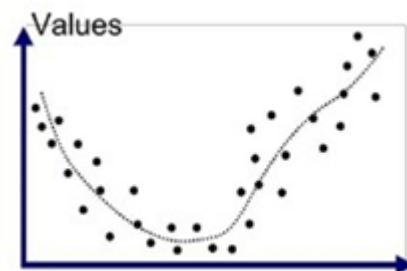
- A la hora de entrenar un modelo, existe un conflicto entre optimización y generalización
- Optimización es el proceso de ajuste de un modelo para conseguir el mejor rendimiento posible de los datos de entrenamiento (proceso de aprendizaje).
- Generalización es cómo de bien se comporta el modelos ante datos no procesados aún. El objetivo es conseguir la mejor capacidad de generalización.
- El modelo comienza a estar sobre-ajustado ó “overfitted” cuando ha aprendido tan bien los datos de entrenamiento, que memoriza pautas que son demasiado específicas de los datos de entrenamiento e irrelevantes para los nuevos datos.



Underfitting

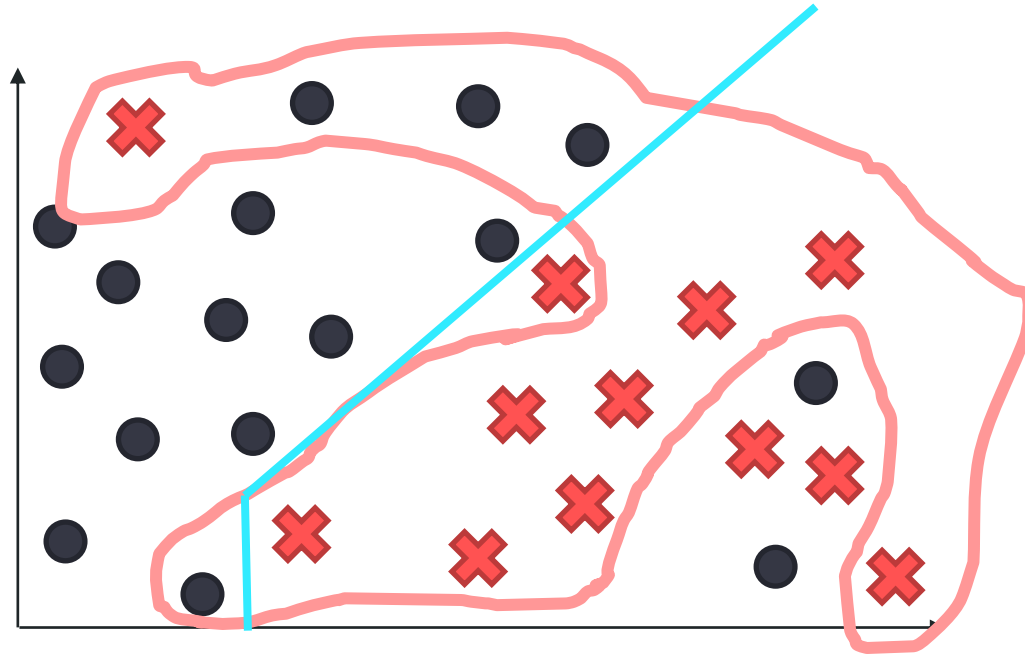


Overfitting



Buen ajuste

# Overfitting y Underfitting

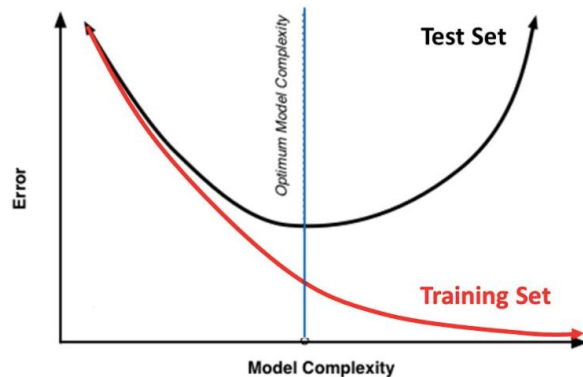


# Overfitting y Underfitting

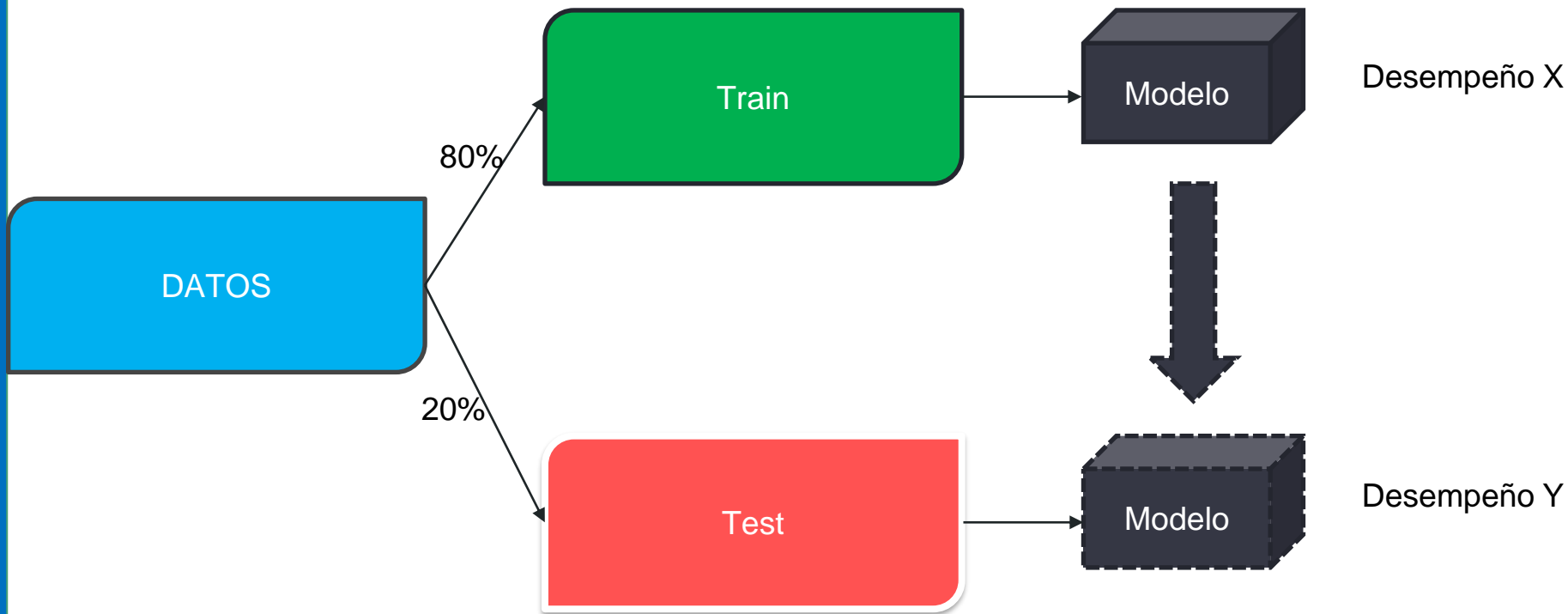
Formas de evitar el “overfitting”:

- **Obtener más datos** es normalmente la mejor solución, un modelo entrenado con más datos generalizará mejor de forma natural.
- La **regularización** es diferente en función del algoritmo y consiste en simplificarlo para evitar que “memorice” los datos de entrenamiento

Training Vs. Test Set Error



# Evaluación de modelos



- Si  $Y \simeq X \Rightarrow$  Buena generalización del modelo
- Si  $Y \ll X \Rightarrow$  Overfitting

# ¡Gracias!

Contacto: Rafael Zambrano

[rafazamb@gmail.com](mailto:rafazamb@gmail.com)