

Guía de lectura

Introduction to Statistical Learning. Springer, 2013

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

Desarrolle los siguientes puntos (**Quiz al comienzo de la sesión 3**) basándose en la lectura. Se puede apoyar en fuentes adicionales (artículos, libros, internet).

Capítulo 2, puntos 2.1 y 2.2

- Explique las diferencias entre Machine Learning y Statistical Learning, basándose en los enfoques de predicción e inferencia y las suposiciones de los modelos.
- ¿Cuál es la diferencia entre modelos paramétricos y no paramétricos?
- ¿Por qué escoger modelos predictivos restrictivos cuando es posible escoger modelos flexibles? Explicar (sin fórmulas matemáticas) el compromiso entre los errores de sesgo (bias) y varianza y su relación con el overfitting.
- ¿Cuáles son las relaciones entre los modelos supervisados, no supervisados, de regresión, y de clasificación?
- ¿Cómo se mide la calidad de un modelo?
- Describa brevemente el funcionamiento del método K-NN, tanto para regresión como para clasificación.

Capítulo 4, puntos 4.1, 4.2 y 4.3

- ¿Por qué un modelo de regresión lineal no se puede utilizar para un problema de clasificación de una variable objetivo de 2 clases?
- Describa brevemente el funcionamiento un modelo de regresión logística.
- ¿Cómo hacemos para utilizar regresión logística en el caso de tener una variable objetivo con más de 2 clases?

Para discutir en clase:

- ¿De qué modelos se habla en el texto explícitamente, y cuáles de ellos uds pueden explicar?
- ¿Qué conceptos se utilizaron y explicaron en la lectura? ¿Qué conceptos se utilizaron en la lectura y no fueron explicados en ella?
- Fenómeno de *confounding variables*.