

EST-25134: Aprendizaje Estadístico

Profesor: Alfredo Garbuno Iñigo — Primavera, 2022 — Árboles de decisión.

Objetivo: Que veremos.

Lectura recomendada: Capítulo 8 de [1].

1. INTRODUCCIÓN

En esta sección estudiaremos modelos basados en **árboles de decisión**. Los cuales estratifican el espacio de los atributos en regiones sencillas para efectuar predicciones. Reciben su nombre pues la segmentación que se utiliza para realizar predicciones es una secuencia de decisiones binarias.

1.1. Ejemplo: Baseball

Consideremos los datos de salarios de jugadores profesionales de *baseball*. Lo que queremos es predecir el Sueldo en funciones de las características de carrera de cada jugador.

| | Hits | Years | Salary | |
|---|------|-------|--------|-----|
| 1 | 1 | 81 | 14 | 475 |
| 2 | 2 | 130 | 3 | 480 |
| 3 | 3 | 141 | 11 | 500 |
| 4 | 4 | 87 | 2 | 92 |
| 5 | 5 | 169 | 11 | 750 |
| 6 | 6 | 37 | 2 | 70 |

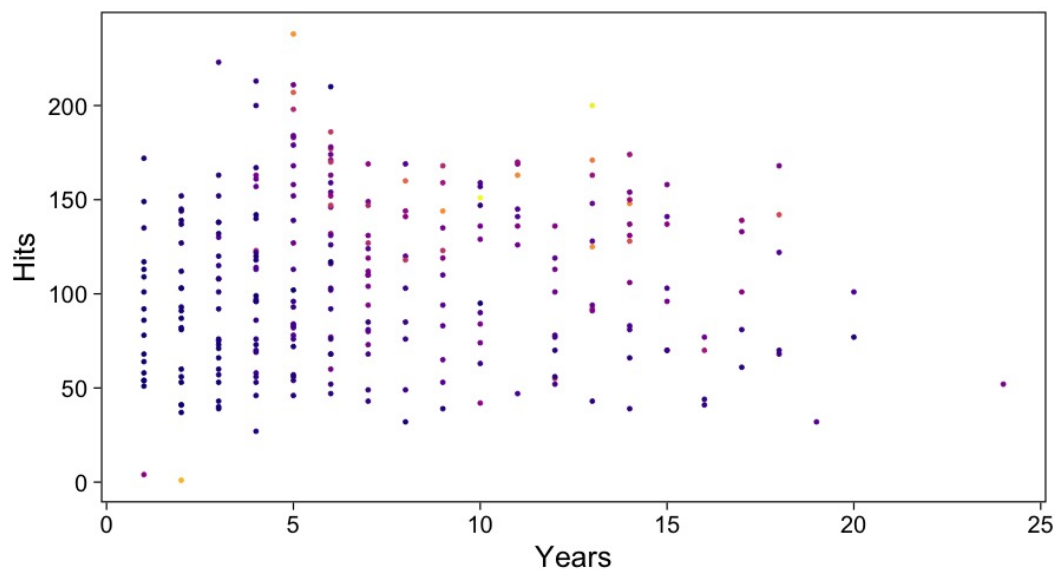


FIGURA 1. Salario codificado por color: salarios bajos (azul, morado) y salarios altos (naranja, amarillo).

```
1 Salary
2   226 when Years < 4.5
3   465 when Years ≥ 4.5 & Hits < 118
4   949 when Years ≥ 4.5 & Hits ≥ 118
```

LISTING 1. Segmentacion de datos utilizando un árbol de decisión.

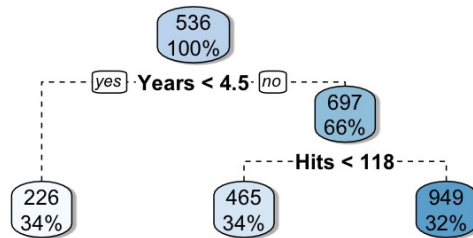


FIGURA 2. Representación grafica de un árbol de decisión.

REFERENCIAS

- [1] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer US, New York, NY, 2021. ISBN 978-1-07-161417-4 978-1-07-161418-1. . [1](#)