

Árboles de decisión.

Datasets con pocos atributos! Pero pueden tener fronteras de clasificación complejas

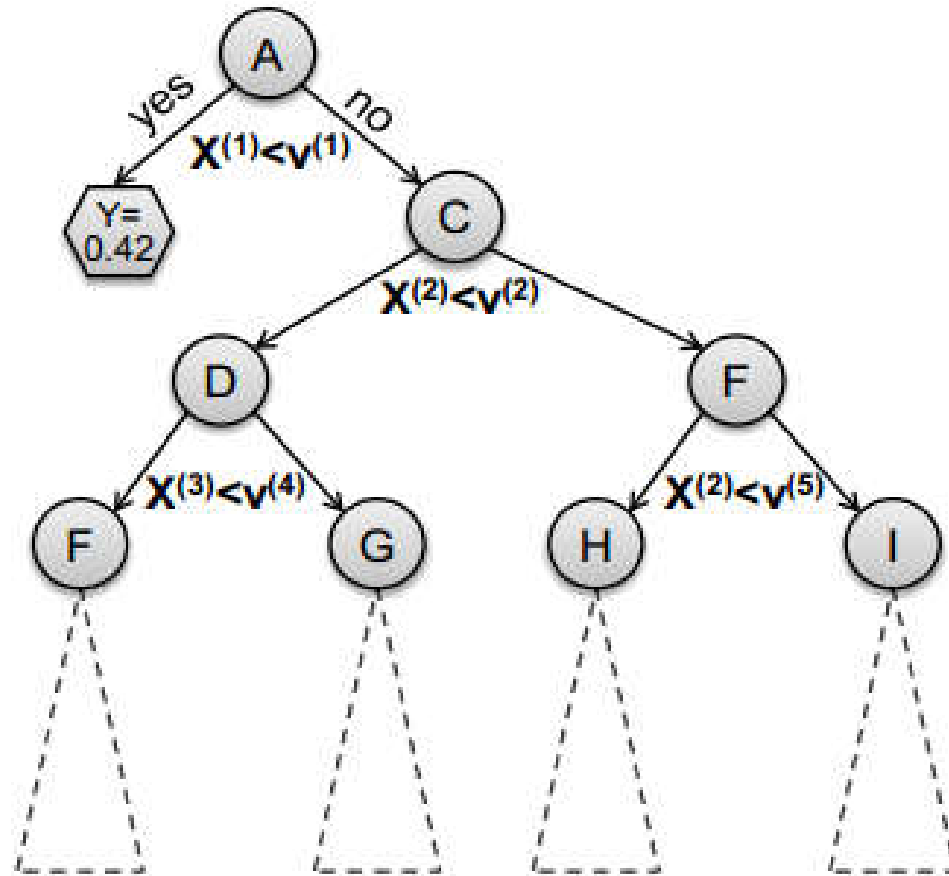
$x \longrightarrow y$

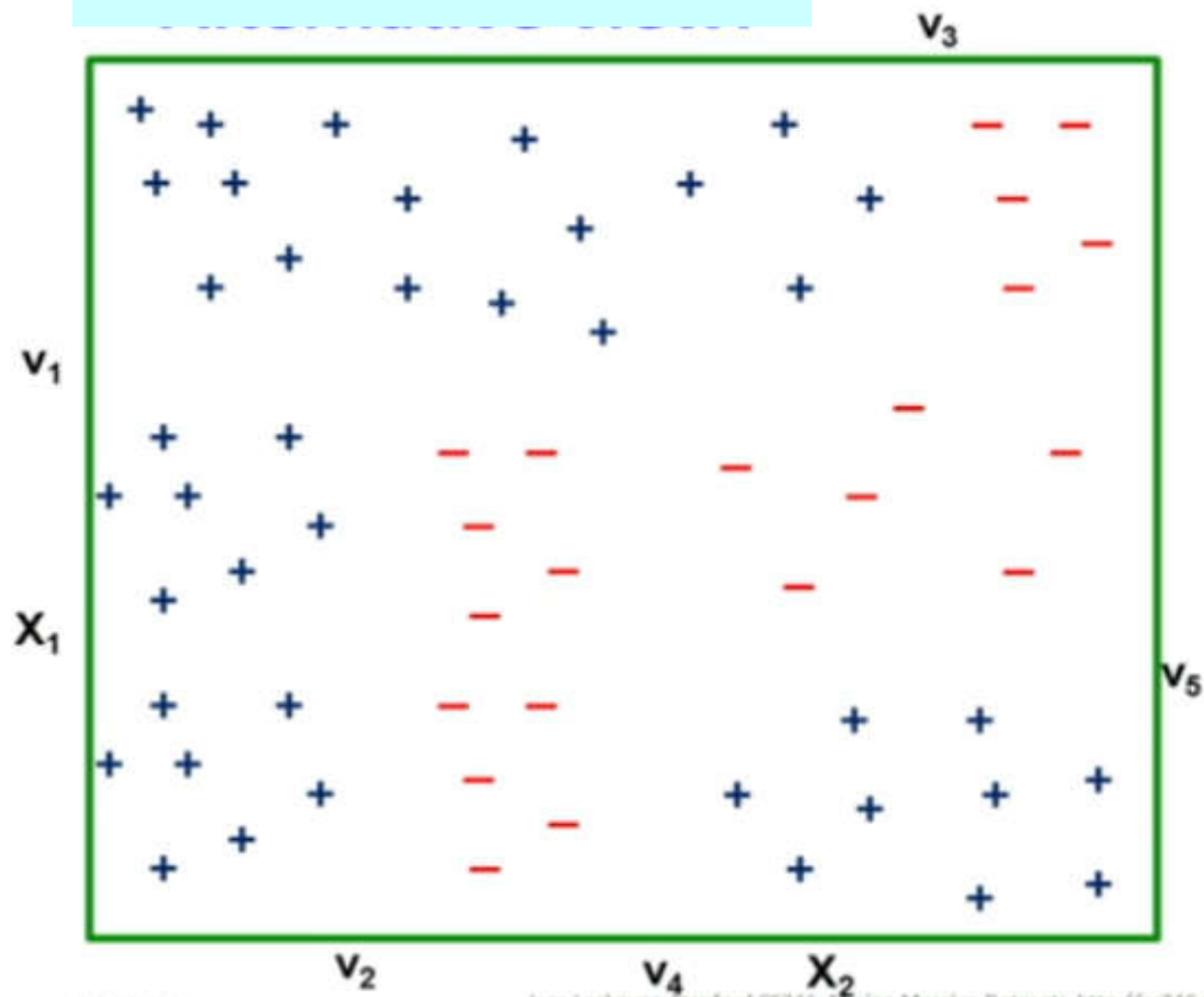
Tipos:

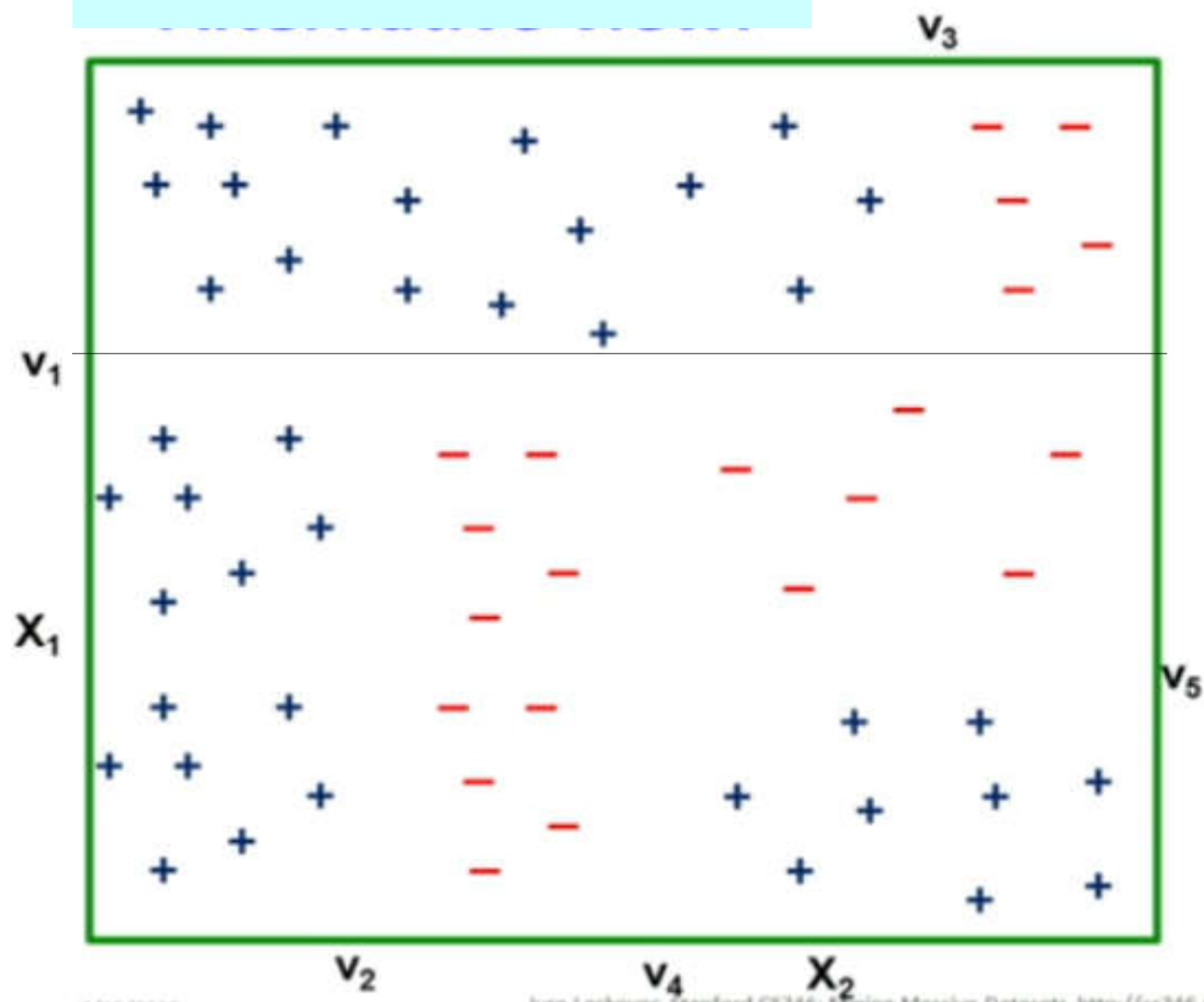
- Regresión
- Clasificación

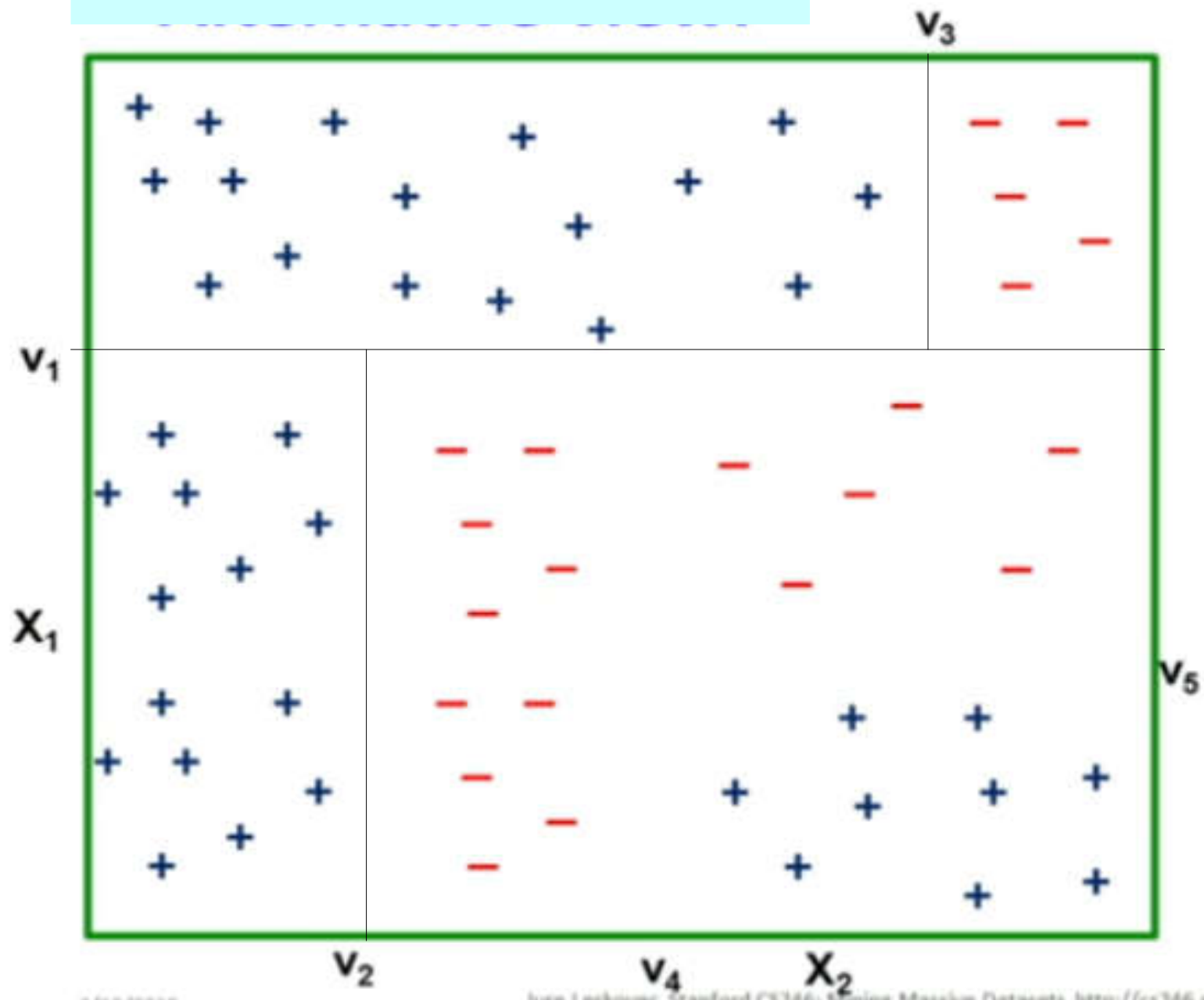
Composición:

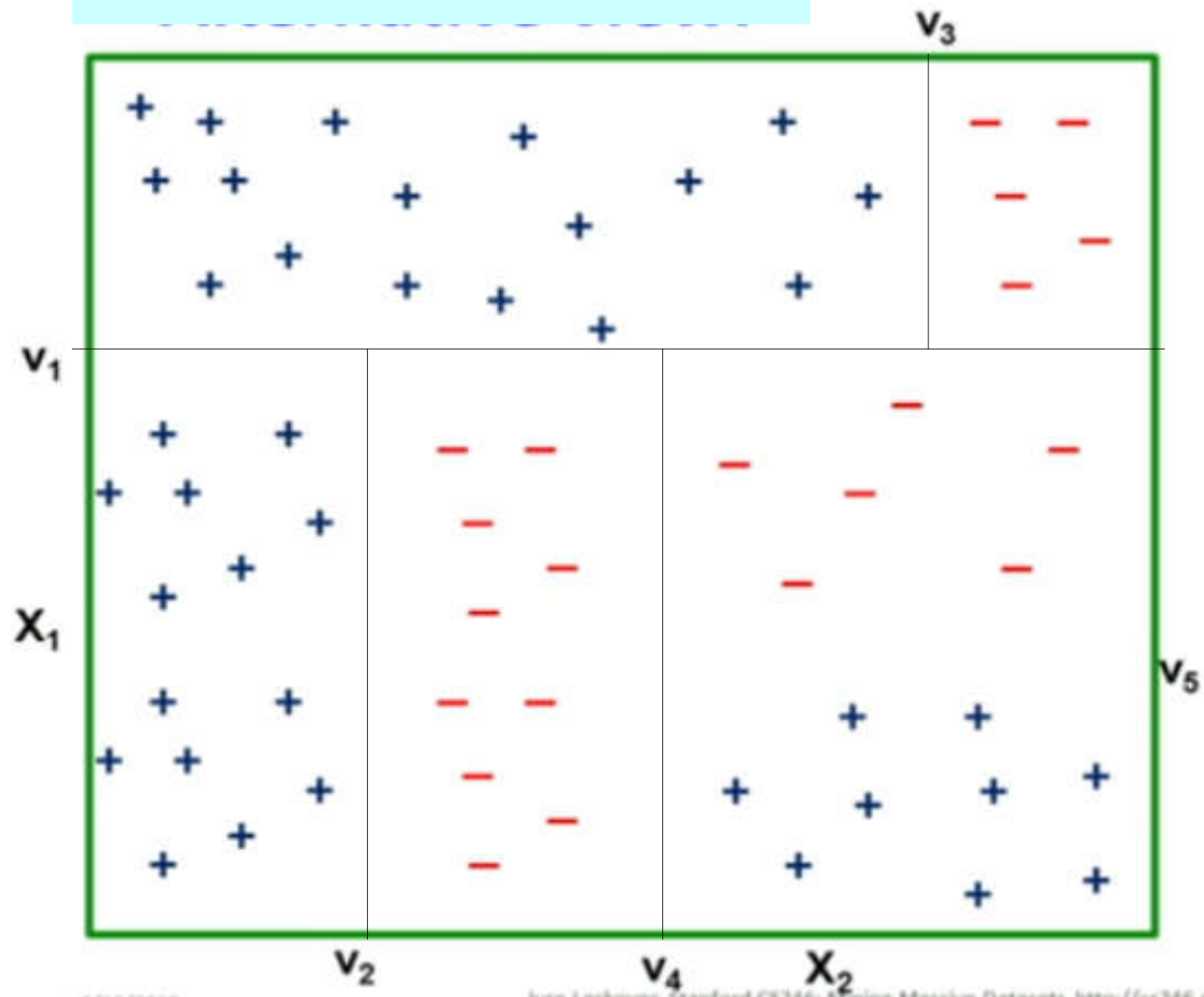
- Nodos: bifurcación
- Hojas: final (predicción)

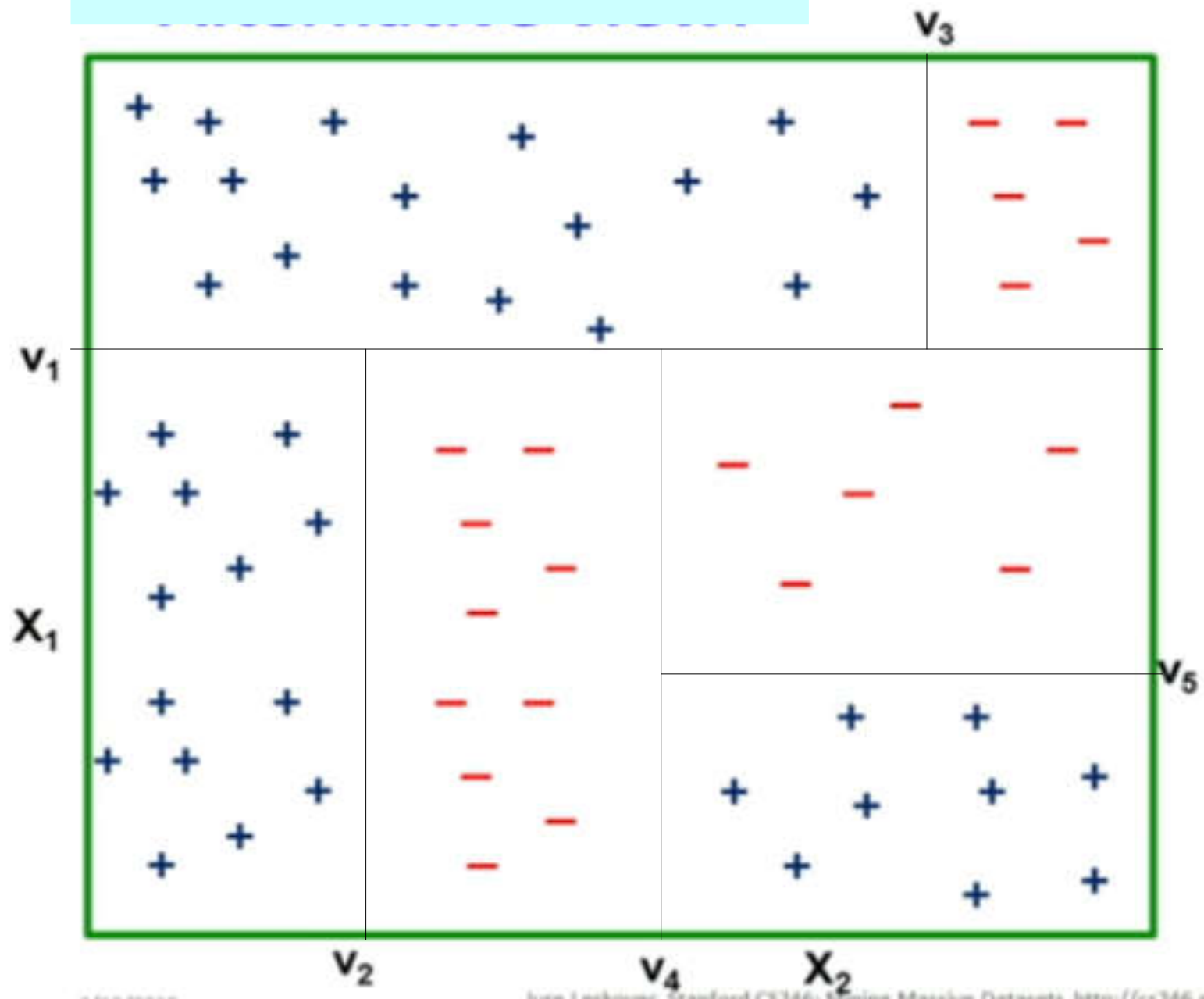


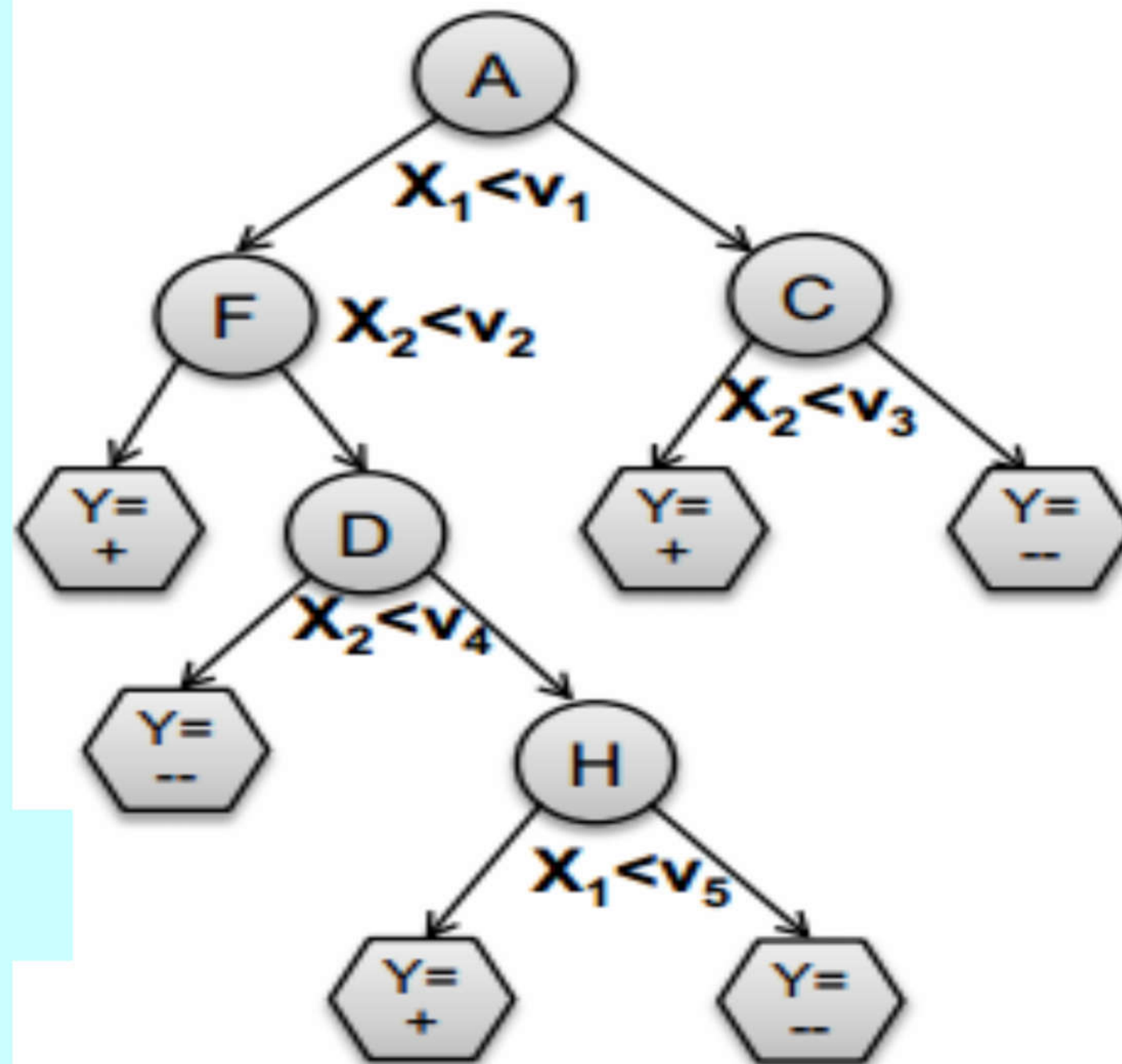












Criterios de parada:

- **Pureza**: La varianza de los elementos de la hoja es menor que un umbral de varianza fijado-
- **Número de elementos** por debajo de cierta cantidad determinada (evitar overfitting).

Criterios de predicción:

- **Regression:**

- Predict average y_i of the examples in the leaf
- Build a linear regression model on the examples in the leaf

- **Classification:**

- Predict most common y_i of the examples in the leaf