Árboles y Random Forest

Árboles y Random Forest

- Árboles
  - Árboles de Regresión
  - Árboles de Clasificación
- Ventajas y desventajas
- Corrección de overfitting
  - Ajuste de Hiperparámetros
  - Pruning
- Equilibrio Sesgo-Varianza

Árboles de Clasificación y Regresión

- **Árboles de Regresión**

  La variable respuesta es numérica

- **Árboles de Clasificación**

  La variable respuesta es nominal (2 o mas estados)

Árboles de prediccion

Algoritmo de partición binaria

1. Para cada variable predictora X:
   1.1 **Puntos de corte**: Su determinan todos los puntos posibles en los que podemos dividir la variable predictora X_j.
   1.2. **Medición de la variabilidad**: Medimos la reducción en la variabilidad para cada partición de la variable respuesta Y.
   1.3. **Corte óptimo**: Se determina el corte que mas reduce la variabilidad.
2. Se selecciona aquella variable X_j con su respectivo punto de corte S que produce la mayor reducción de la varianza. Se ingresa al modelo generando la partición.
3. Se repiten los pasos 1 y 2 hasta que se cumplan las condiciones de corte establecidas.

Árboles de prediccion

Algoritmo de partición binaria

- Si los **predictores continuos**, se ordenan de menor a mayor sus valores, el punto intermedio entre cada par de valores se testea como posible punto de corte.

- Si la **variable es ordinal,** se consideran todos los grupos posibles manteniendo el orden de las categorías.

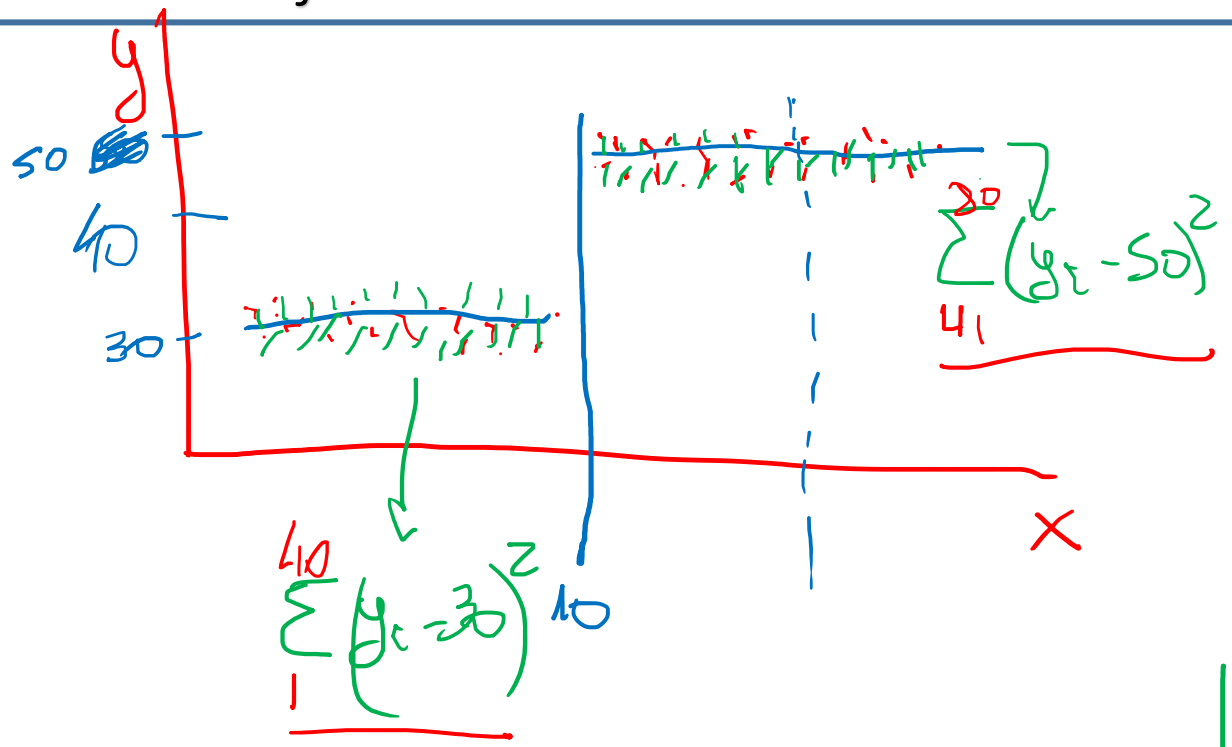
- Si la **variable es nominal,** se deben considerar todos los posibles agrupamientos en 2 grupos y evaluar la función objetivo en cada uno de ellos.

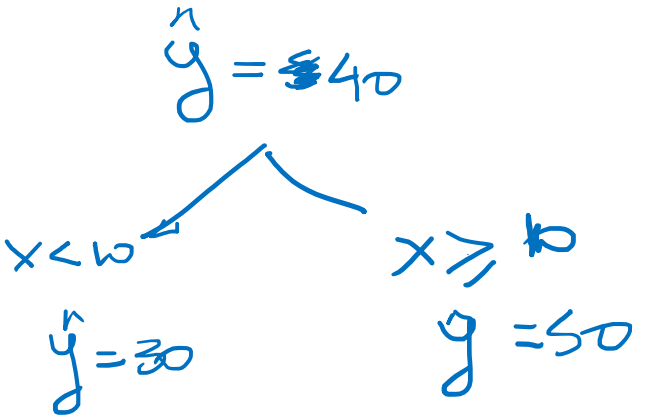
Árboles y Random Forest

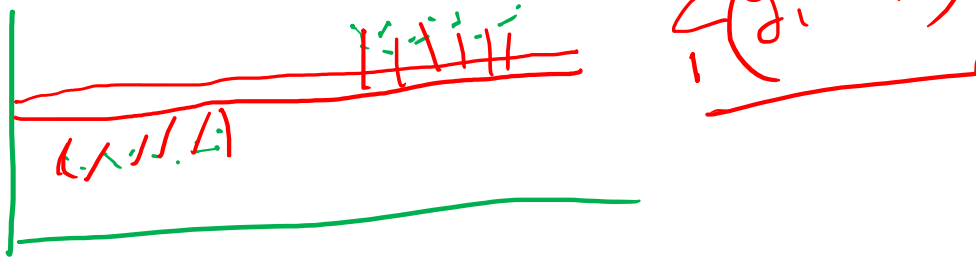
- Árboles
  - Árboles de Regresión
  - Árboles de clasificación
- Ventajas y desventajas
- Corrección de overfitting
  - Ajuste de Hiperparámetros
  - Pruning
- Equilibrio Sesgo-Varianza

Árboles de Regresión

RSS: Función objetivo a minimizar







Árboles de Regresión

RSS: Función objetivo a minimizar

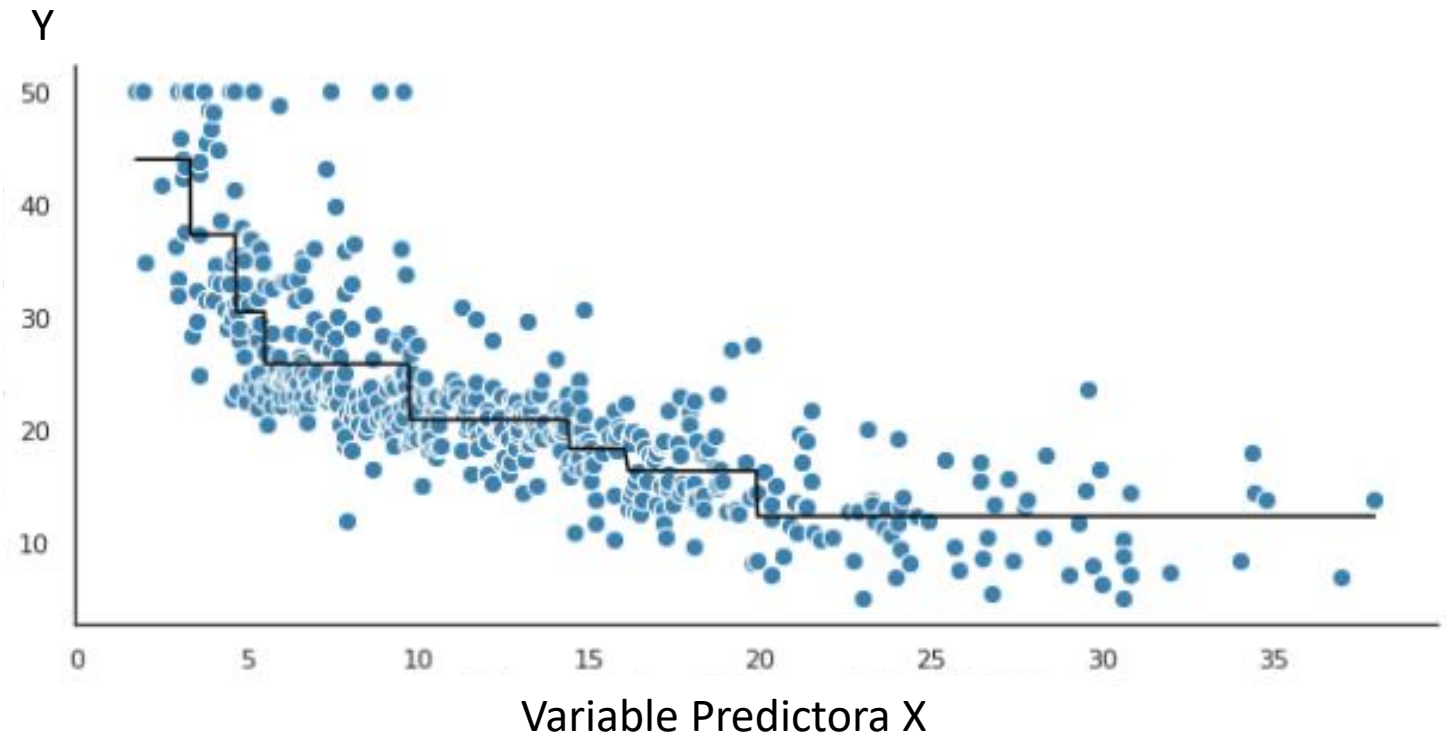
Se calcula el RSS total que se consigue con cada partición evaluada

$$RSS_1 + RSS_2 = \sum_{i:x_i \in R_1} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2} (y_i - \hat{y}_{R_2})^2$$

donde el primer término es el RSS de la región 1 y el segundo término es el RSS de la región 2, siendo cada una de las regiones el resultado de separar las observaciones acorde al predictor j y valor s.

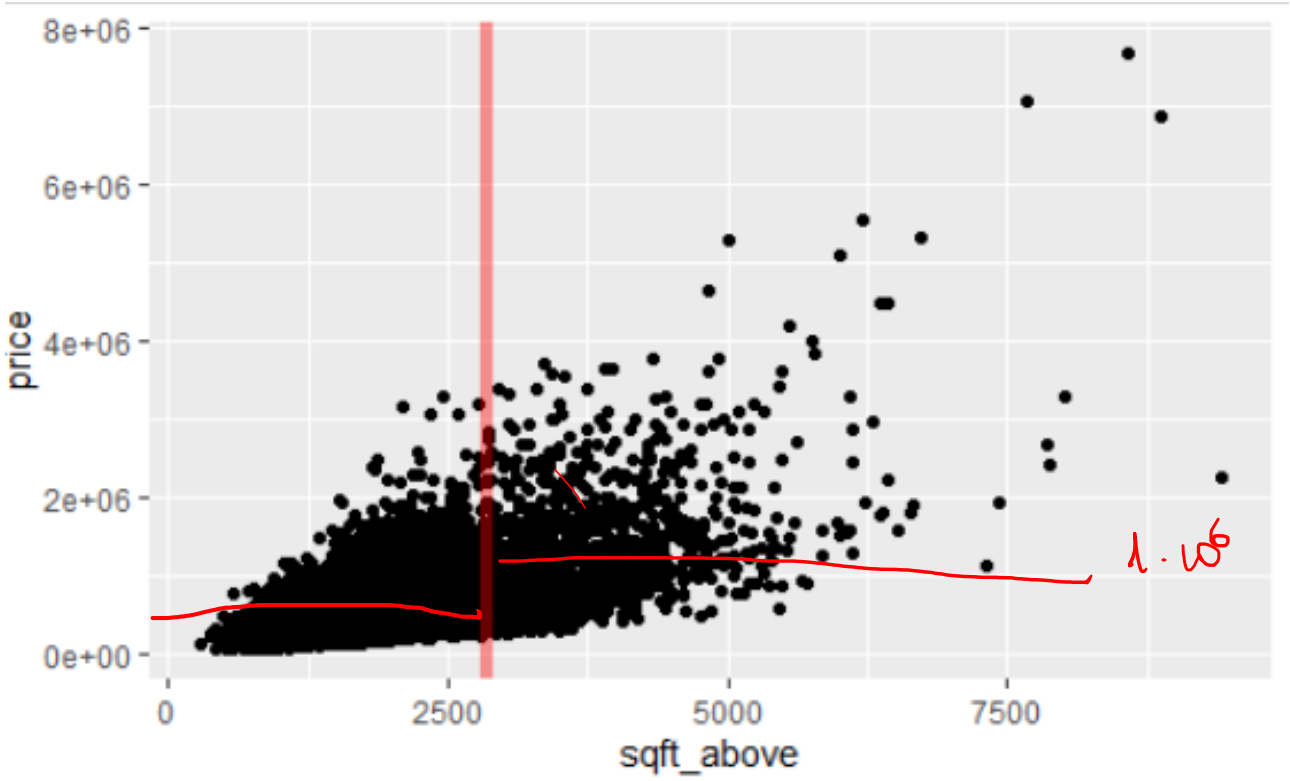
Árboles de Regresión

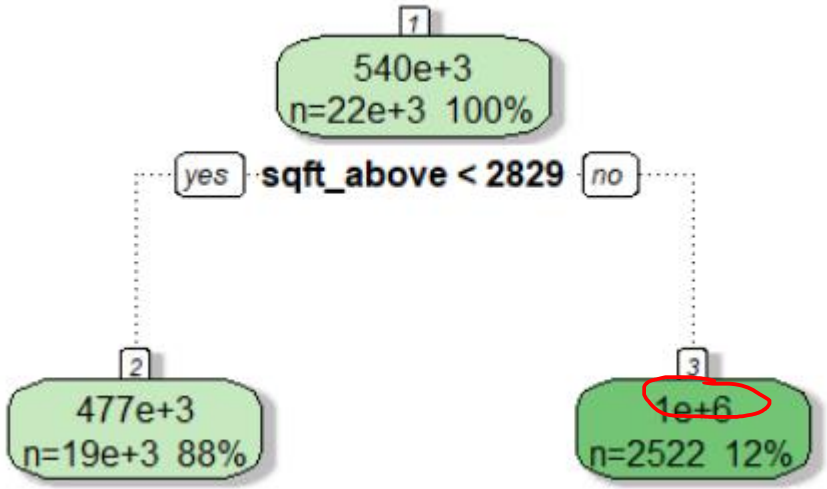
RSS: Función objetivo a minimizar



Árboles de Regresión

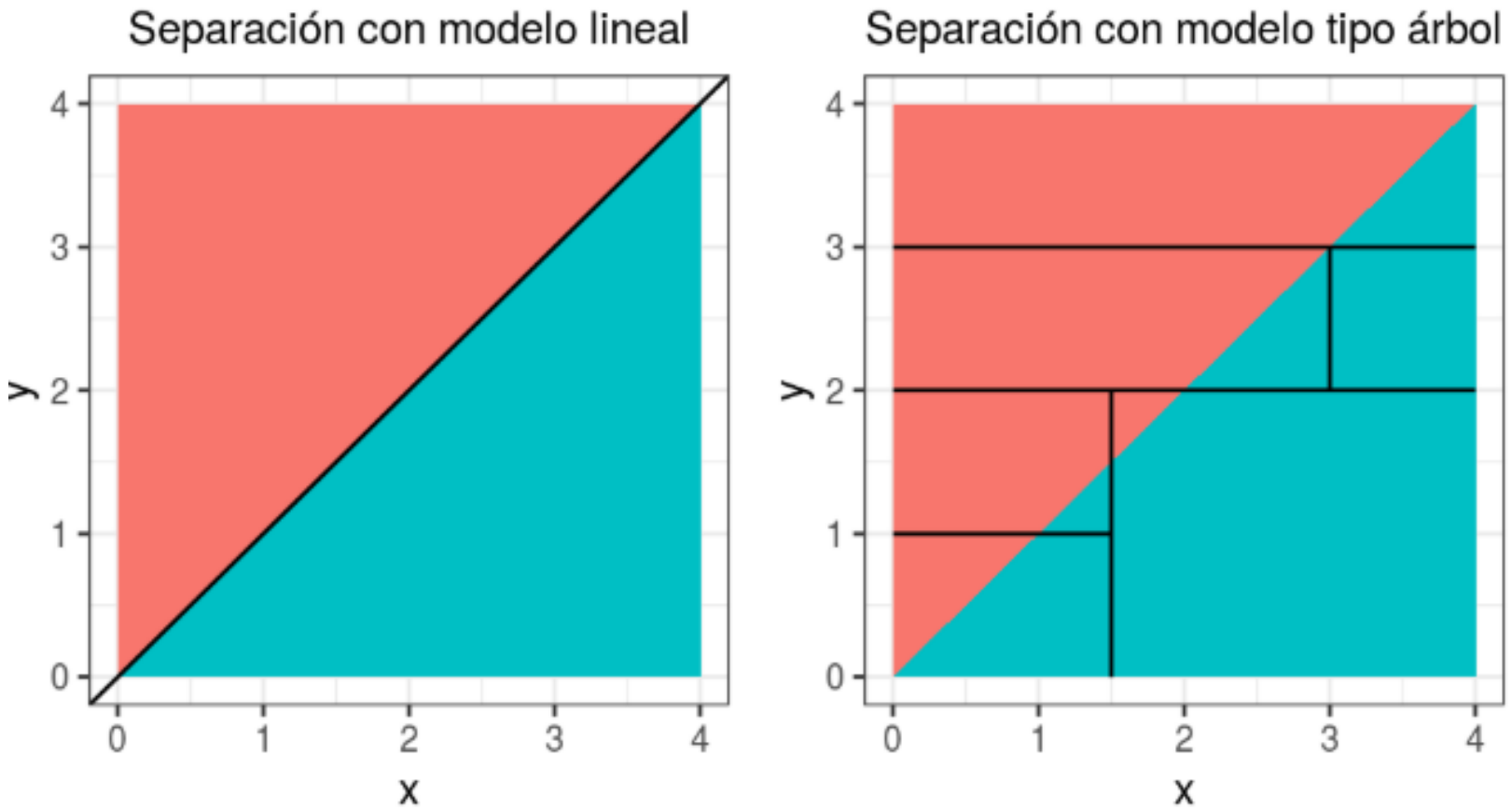
Ejemplo datos Propiedades





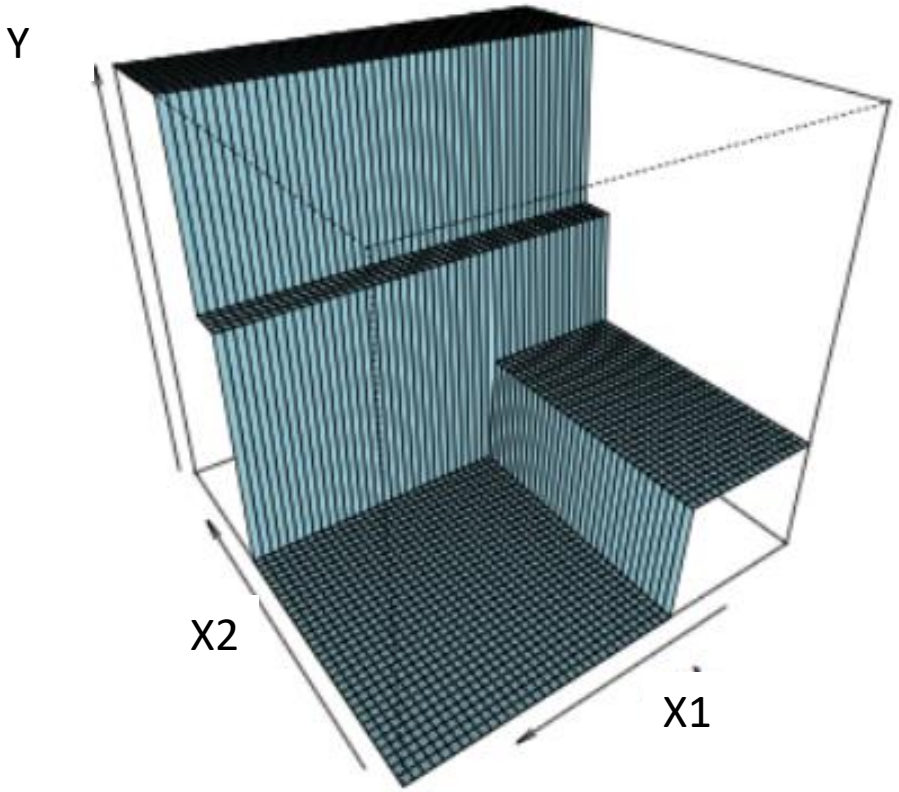
Árboles de Regresión

Clasificación de árbol vs Clasificador lineal



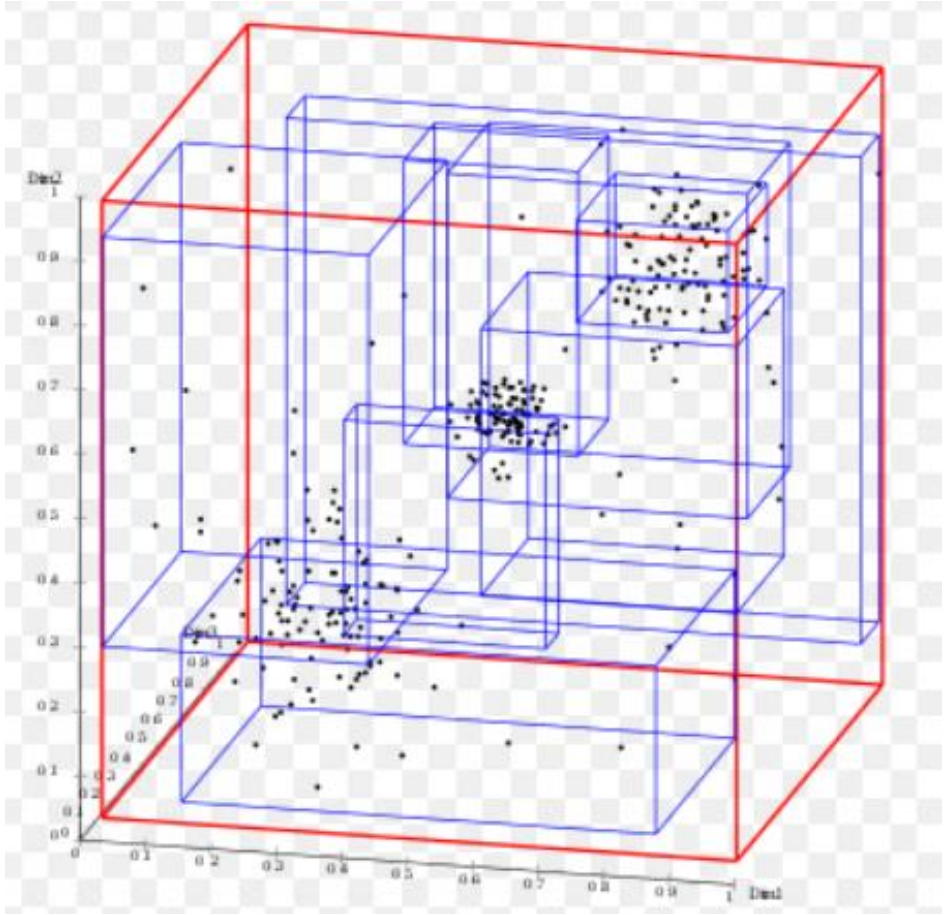
Árboles de Regresión

Clasificación de árbol 2 variables explicativas



Árboles de Regresión

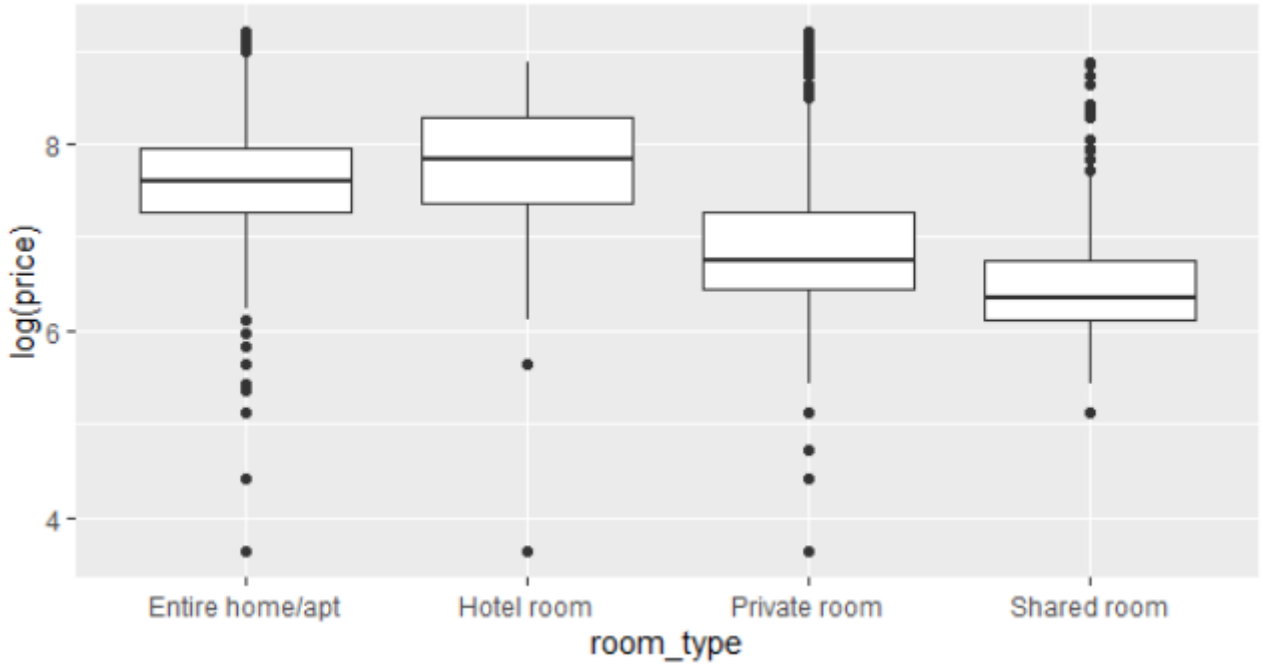
Clasificación de árbol 3 variables explicativas

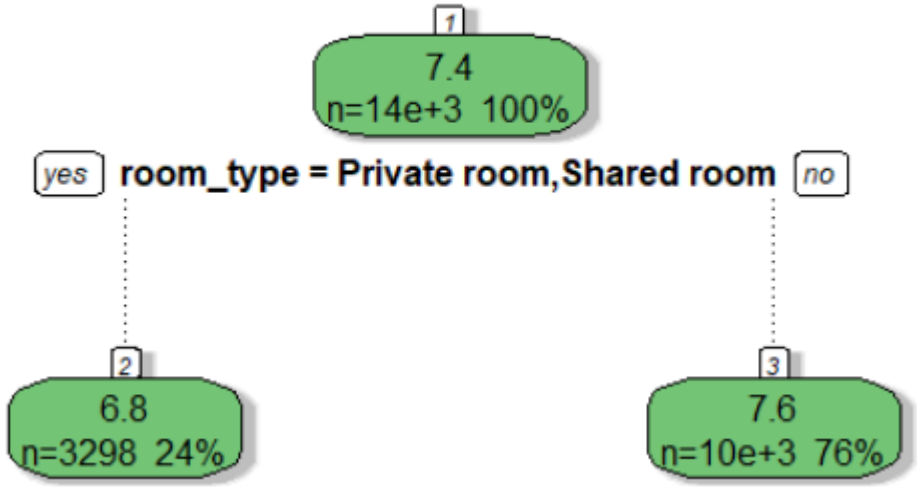


Árboles de clasificación

Predictores no cuantitativos

Debe prestarse especial atención la codificación como factores en las variables explicativas





Árboles y Random Forest

- Árboles
  - Árboles de Regresión
  - Árboles de clasificación
- Ventajas y desventajas
- Corrección de overfitting
  - Ajuste de Hiperparámetros
  - Pruning
- Equilibrio Sesgo-Varianza

Árboles de clasificación

Funciones objetivo a minimizar

Índice de Gini

Es una medida de la varianza total en el conjunto de las *K* clases del nodo *m*. Se considera una medida de pureza del nodo.

$$G_m = \sum_{k=1}^{K} \hat{p}_{mk} \left(1 - \hat{p}_{mk}\right)$$

Árboles de clasificación

Funciones objetivo a minimizar

Ganancia de Información: Entropía cruzada

La entropía es otra forma de cuantificar el desorden de un sistema. En el caso de los nodos, el desorden se corresponde con la impureza. Si un nodo es puro, contiene únicamente observaciones de una clase, su entropía es cero. Por el contrario, si la frecuencia de cada clase es la misma, el valor de la entropía alcanza el valor máximo de 1.

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \; log(\hat{p}_{mk})$$

Árboles y Random Forest

- Árboles
  - Árboles de Regresión
  - Árboles de clasificación
- Ventajas y desventajas
- Corrección de overfitting
  - Ajuste de Hiperparámetros
  - Pruning
- Equilibrio Sesgo-Varianza
  - Bagging
  - Boosting

Árboles de Regresión y Clasificación

Ventajas

- Son fáciles de interpretar. Se pueden representar gráficamente.
- Se pueden usar predictores cuantitativos como cualitativos. No requiere codificación en Dummies.
- Es un método no paramétrico. No requiere supuestos distribucionales.
- Requieren mínimo preprocesamiento.
- Son robustos a la presencia de outliers.
- Seleccionan los mejores predictores en forma automática.

Árboles de Regresión y Clasificación

Desventajas

- El algoritmo *recursive binary splitting* no es óptimo. El orden en que se seleccionan las variables sigue un criterio heurístico que puede condicionar los cortes siguientes.
- Los árboles tienen una tendencia al Sobreajuste (Overfitting). Para resolverlo se utilizan dos técnicas:
  – Control de los parámetros del árbol
  – Pruning (Poda)
- Como todo modelo estadístico, los árboles están sujetos al problema de sesgo-varianza. Para resolverlo existen dos métodos:
  – Bagging
  – Boosting

Árboles y Random Forest

- Árboles
  - Árboles de Regresión
  - Árboles de clasificación
- Ventajas y desventajas
- Corrección de overfitting
  - Ajuste de Hiperparámetros
  - Pruning
- Equilibrio Sesgo-Varianza

Soluciones al Overfitting

Ajuste de Hiperparámetros

- Observaciones mínimas para división
- Observaciones mínimas de nodo terminal
- Profundidad máxima del árbol
- Número máximo de nodos terminales
- Reducción mínima de error
- Hay mas…

Soluciones al Overfitting

Pruning

Cost complexity pruning

Es análogo al método empleado en ridge regression o lasso.

En este caso, se busca el sub-árbol T que minimiza la ecuación:

$$\sum_{j=1}^{|T|} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 + \alpha |T|$$

donde $|T|$ es el número de nodos terminales del árbol.

Árboles y Random Forest

- Árboles
  - Árboles de Regresión
  - Árboles de clasificación
- Ventajas y desventajas
- Corrección de overfitting
  - Ajuste de Hiperparámetros
  - Pruning
- Equilibrio Sesgo-Varianza

Equilibrio Sesgo-Varianza

- Bagging
  - Bagging
  - Bagging + Random Forest
  - Bagging + Extremely Randomized Trees

- Boosting
  - AdaBoost
  - Gradient Boosting
  - Stochastic Gradient Boosting

Bagging ("Bootstrap aggregation")

Se emplea el muestreo repetido (bootstrapping) con el fin de reducir la varianza.

1. Se generan T *pseudo-training sets* mediante bootstrapping a partir de la muestra de entrenamiento original.
2. Se entrena un árbol con cada una de las T muestras del paso 1. Cada árbol se crea sin restricciones ni pruning, por lo que tiene varianza alta pero poco sesgo. Ejemplo: la única regla de parada puede ser el número mínimo de observaciones que deben tener los nodos terminales.
3. Para cada nueva observación, obtener la predicción de cada uno de los T árboles. El valor final de la predicción se obtiene como la media de las T predicciones en el caso de variables cuantitativas y como la clase predicha más frecuente (modo) para las variables cualitativas.

Bagging + Random Forest

Si los resultados de los modelos de los distintos árboles se encuentran correlacionados, es difícil reducir la varianza con el procedimiento de bagging.

Random Forest

Para incorrelacionar los resultados, cada vez que se analiza una partición para una variable respuesta, se toma una muestra aleatoria de m variables predictoras de un total de p variables.

Bagging + Extremely Randomized Trees

Los métodos de *Extremely Randomized Trees* llevan la decorrelación de los árboles un paso más allá que *Random Forest*. En cada división de los nodos, solo evalúa un subconjunto aleatorio de los predictores disponibles y, además, dentro de cada predictor seleccionado solo se evalúa un subconjunto aleatorio de los posibles puntos de corte. En determinados escenarios, este método consigue reducir un poco más la varianza. El paquete ExtraTrees implementa algoritmos de este tipo.

Equilibrio Sesgo-Varianza

- Bagging
  - Bagging
  - Bagging + Random Forest
  - Bagging + Extremely Randomized Trees

- Boosting
  - AdaBoost
  - Gradient Boosting
  - Stochastic Gradient Boosting

XGBoost eXtreme Gradient Boosting