{desafío} latam_

Árboles de Decisión _

Motivación

¿Qué son?

Métodos que buscan particionar el espacio de atributos en una serie de rectángulos y posteriormente se implementa un modelo simple (o estadístico de representación) [Definición de Hastie et al. 2009].

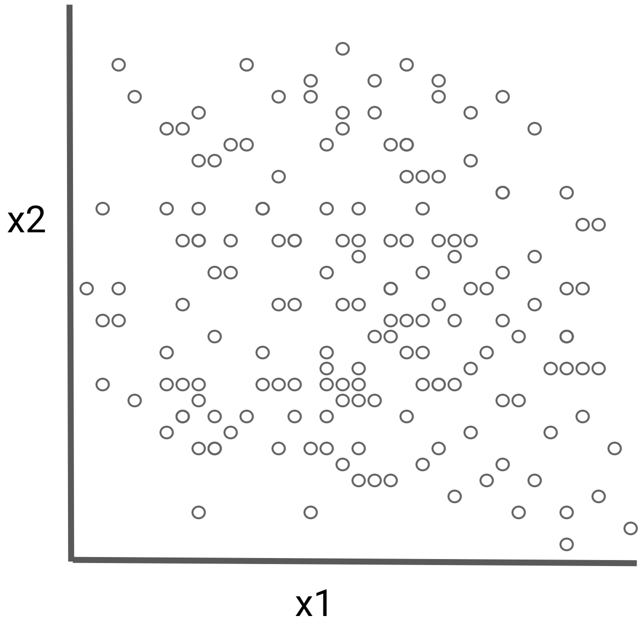
Virtudes:

- Entendibles (a diferencia de los modelos de caja negra como SVM y EM). Por lo general tienen buenos resultados.
- Son no-paramétricos: Encontrarán una solución o representación al problema sin la necesidad de definir componentes estocásticos/sistemáticos (como una regresión lineal).

Defectos:

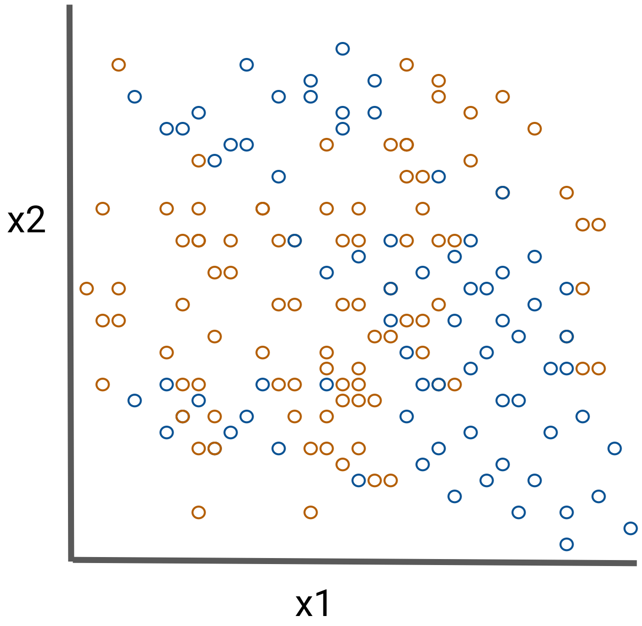
- Estos buenos resultados tienden a generarse en el training, lo cual induce el overfit.

El proceso de decisión

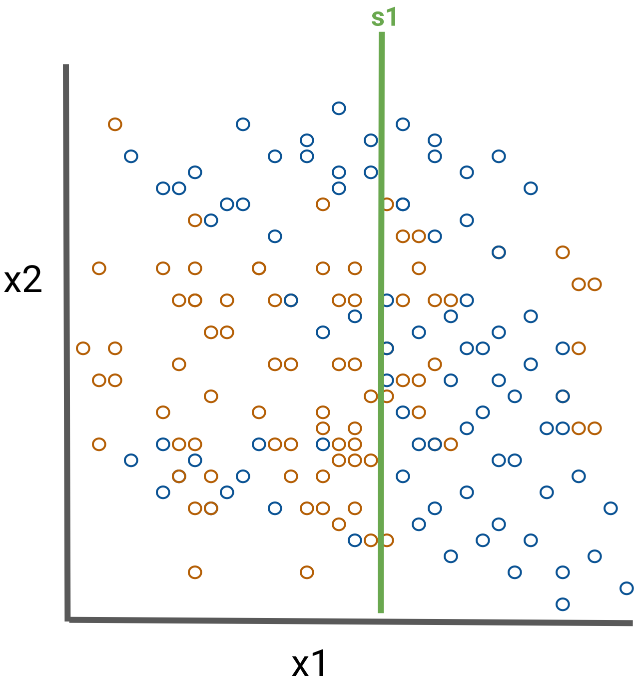


{desafío} latam_

El proceso de decisión

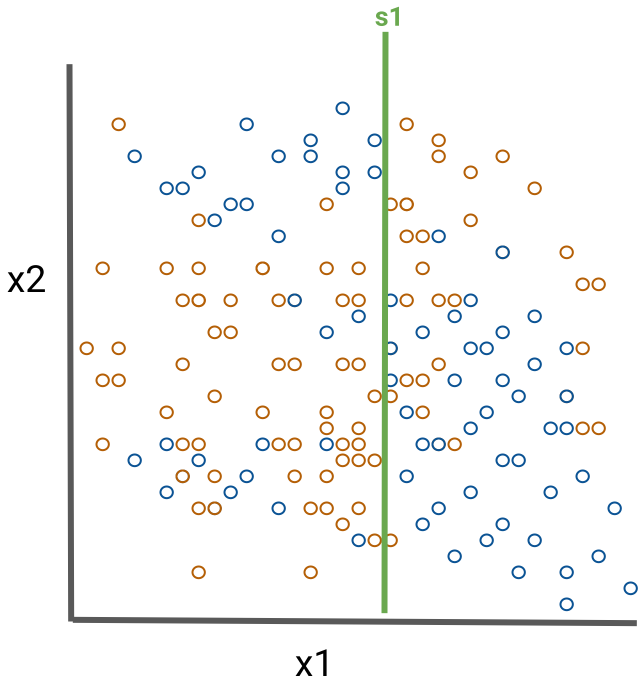


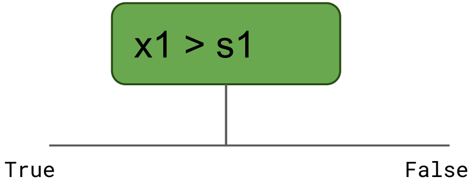
El proceso de decisión



x1 > s1

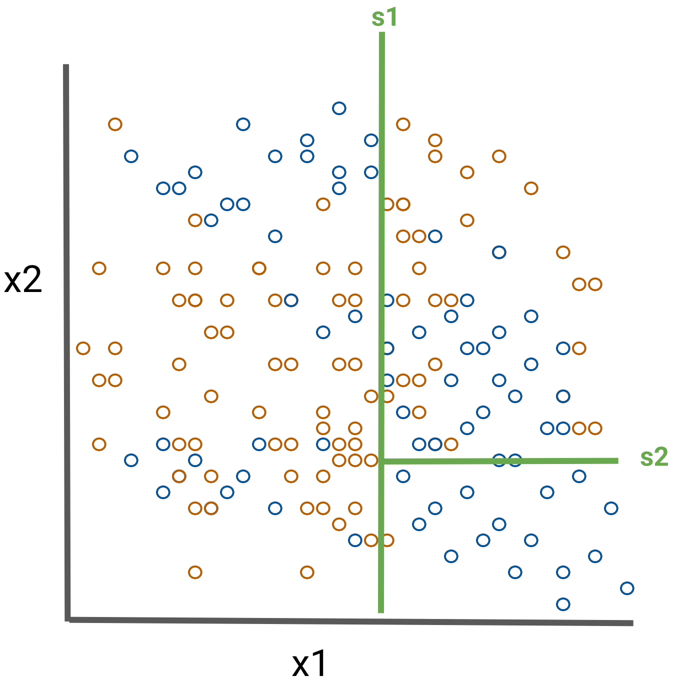
El proceso de decisión

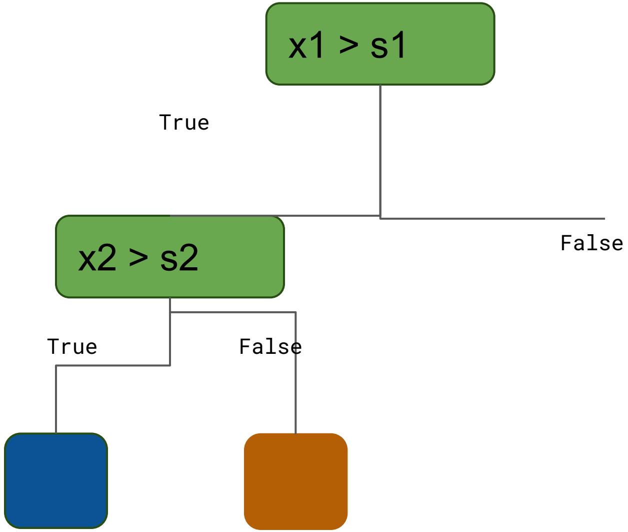




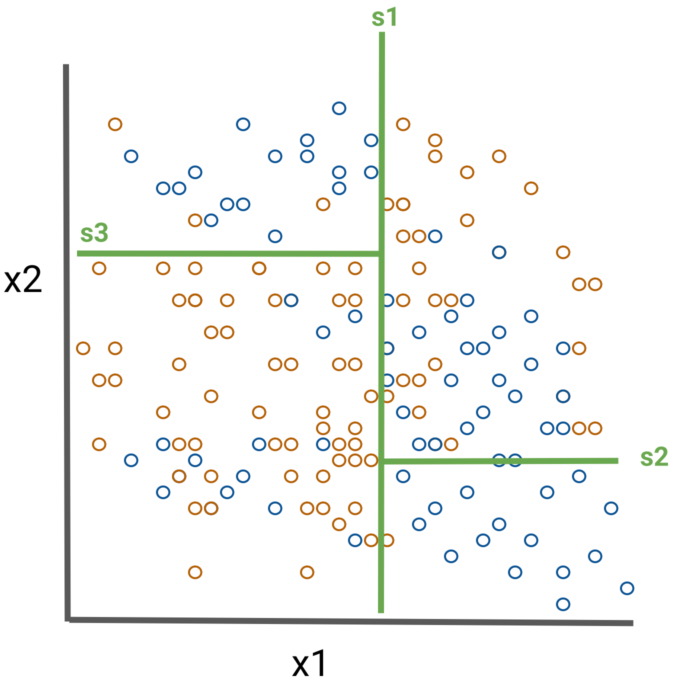
{desafío}
latam_

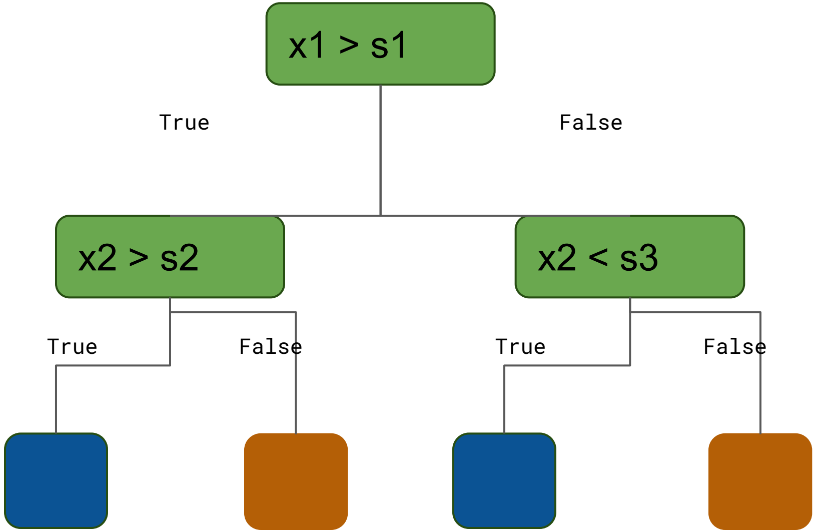
El proceso de decisión





El proceso de decisión





Rudimentos

Preliminares

$$\hat{f}(x) = \sum_{m=1}^{s} c_m \mathbb{I}\big((X_1, X_2) \in \mathbb{R}_m\big)$$

Nuestro objetivo es desarrollar una función de regresión (con constantes) que resuma la subdivisión.

Preliminares

$$\hat{f}(x) = \sum_{m=1}^{s} c_m \mathbb{I}((X_1, X_2) \in \mathbb{R}_m)$$

C representa la
cantidad de constantes
estimables en el árbol

Preliminares

$$\hat{f}(x) = \sum_{m=1}^{s} c_m \mathbb{I}((X_1, X_2) \in \mathbb{R}_m)$$

R representa las hojas condicionadas al subespacio de valores.

{desafío} latam_

Preliminares

$$\hat{f}(x) = \sum_{m=1}^{s} c_m \mathbb{I}((X_1, X_2) \in \mathbb{R}_m)$$

La sumatoria representa todo el proceso de crear un árbol, donde S va a determinar su profundidad.

Elección de una nueva decisión

$$\mathbb{R}_1(j, s) = \{X | X_j \leq s\} \quad \text{ó} \quad \mathbb{R}_2(j, s) = \{X | X_j \geq s\}$$

Dada una función de regresión que establezca una serie de constantes en un modelo de regresión, la decisión se reduce un problema de optimización greedy donde dado una variable j y un punto de corte s, definimos mitades de plano de manera tal que la elección refleja el puntaje.

Elección de una nueva decisión

$$\mathbb{R}_1(j, s) = \{X | X_j \leq s\} \quad \text{ó} \quad \mathbb{R}_2(j, s) = \{X | X_j \geq s\}$$

Dada una función de regresión que establezca una serie de constantes en un modelo de regresión, la decisión se reduce un problema de optimización greedy donde dado una variable j y un punto de corte s, definimos mitades de plano de manera tal que la elección refleja el puntaje.

Elección de una nueva decisión

$$\mathbb{R}_1(j, s) = \{X | X_j \leq s\} \quad \text{ó} \quad \mathbb{R}_2(j, s) = \{X | X_j \geq s\}$$

Dada una función de regresión que establezca una serie de constantes en un modelo de regresión, la decisión se reduce un problema de optimización greedy donde dado una variable j y un punto de corte s, definimos mitades de plano de manera tal que la elección refleja el puntaje.

{desafío} latam_

Elección de una nueva decisión

$$\mathbb{R}_1(j, s) = \{X | X_j \leq s\} \quad \text{ó} \quad \mathbb{R}_2(j, s) = \{X | X_j \geq s\}$$

Dada una función de regresión que establezca una serie de constantes en un modelo de regresión, la decisión se reduce un problema de optimización greedy donde dado una variable j y un punto de corte s, definimos mitades de plano de manera tal que la elección refleja el puntaje.

Elección de una nueva decisión

$$\mathbb{R}_1(j, s) = \{X | X_j \leq s\} \quad \text{ó} \quad \mathbb{R}_2(j, s) = \{X | X_j \geq s\}$$

Dada una función de regresión que establezca una serie de constantes en un modelo de regresión, la decisión se reduce un problema de optimización greedy donde dado una variable j y un punto de corte s, definimos mitades de plano de manera tal que la elección refleja el puntaje.

{desafío}
latam_

Optimización del corte

$$\underset{j,s}{\text{argmin}}\left[\underset{c_1}{\text{argmin}}\sum_{x_i\in\mathbb{R}}(y_i-c_1)^2+\underset{c_2}{\text{argmin}}\sum_{x_i\in\mathbb{R}}(y_i-c_2)^2\right]$$

La elección del corte depende a nivel de atributo y la optimización de cada constante candidata generada

Optimización del corte

$$\underset{j,s}{\text{argmin}}\left[\underset{c_1}{\text{argmin}}\sum_{x_i \in \mathbb{R}}(y_i - c_1)^2 + \underset{c_2}{\text{argmin}}\sum_{x_i \in \mathbb{R}}(y_i - c_2)^2\right]$$

Evaluamos la constante 1 en el subplano de decisión, minimizando la impureza

Optimización del corte

$$\underset{j,s}{\operatorname{argmin}}\left[\underset{c_1}{\operatorname{argmin}} \sum_{x_i \in \mathbb{R}} (y_i - c_1)^2 + \underset{c_2}{\operatorname{argmin}} \sum_{x_i \in \mathbb{R}} (y_i - c_2)^2\right]$$

Evaluamos la constante 2 en el subplano de decisión, minimizando la impureza

{desafío} latam_

Optimización del corte

$$\underset{j,s}{\text{argmin}}\left[\underset{c_1}{\text{argmin}}\sum_{x_i \in \mathbb{R}}(y_i - c_1)^2 + \underset{c_2}{\text{argmin}}\sum_{x_i \in \mathbb{R}}(y_i - c_2)^2\right]$$

Al sumar cada optimización local, evaluamos su comportamiento a nivel de divisor y atributo

Hiperparámetros

Hiperparámetros asociados

Los hiperparámetros en un árbol de decisión buscan controlar la tendencia de crecer de manera irrestricta:

- ¿Hasta qué punto puedo dejar crecer un árbol?
- ¿Cuántos datos son suficientes en cada nodo para particionar o declararlo terminal?
- ¿Cuántos atributos son suficientes para que mi árbol pueda capturar de buena manera el fenómeno?

Hiperparámetros asociados

- **Máximo de Profundidad:**
  Hasta cuándo puede crecer un árbol.

- **Cantidad de atributos:**
  Cuántos atributos debo considerar en un árbol.

- **Mínimo de muestras en un nodo particionable:**
  Con cuántas observaciones podemos seguir subdividiendo.

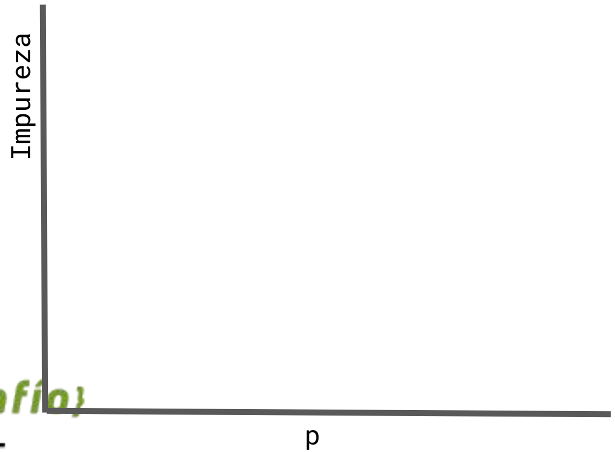
- **Mínimo de muestras en un nodo terminal:**
  Con cuántas observaciones paramos de subdividir.

Importancia Relativa

- Miden la influencia de un atributo en la decisión del árbol.
- Atributos con una mayor carga se posicionarán antes en el proceso de decisión.
- Cargas altas significan que la división fue más pura.
- **Cargas bajas no significan que el atributo sea irrelevante**: Pueden significar que el modelo no evaluó bien en los atributos.

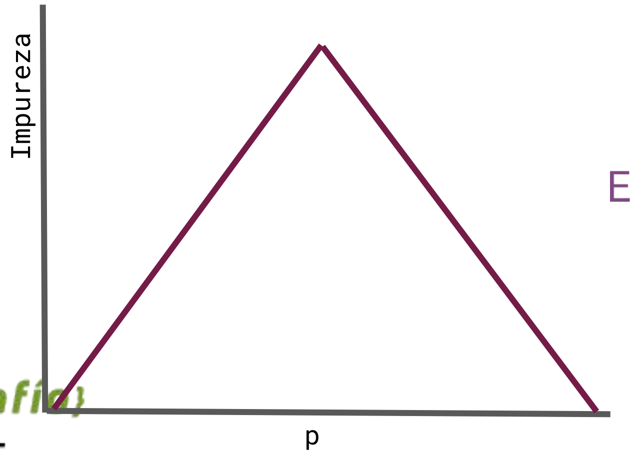
Impurezas en clasificación

- En un árbol de regresión, el problema se reduce a encontrar una combinación de variable y divisor que reduzca el error cuadrático promedio.
- Para el caso de la clasificación, la medida de impureza conlleva evaluar la tasa de impureza en el modelo.



Impurezas en clasificación

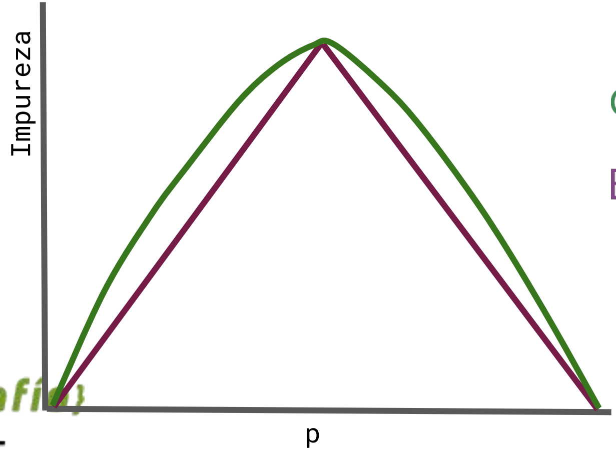
- En un árbol de regresión, el problema se reduce a encontrar una combinación de variable y divisor que reduzca el error cuadrático promedio.
- Para el caso de la clasificación, la medida de impureza conlleva evaluar la tasa de impureza en el modelo.



$$\text{Error de clasificacion} = \frac{1}{N_m}\sum_{i\in\mathbb{R}_m} I(y_i \neq k(m)) = 1 - \hat{p}_{mk(m)}$$

Impurezas en clasificación

- En un árbol de regresión, el problema se reduce a encontrar una combinación de variable y divisor que reduzca el error cuadrático promedio.
- Para el caso de la clasificación, la medida de impureza conlleva evaluar la tasa de impureza en el modelo.

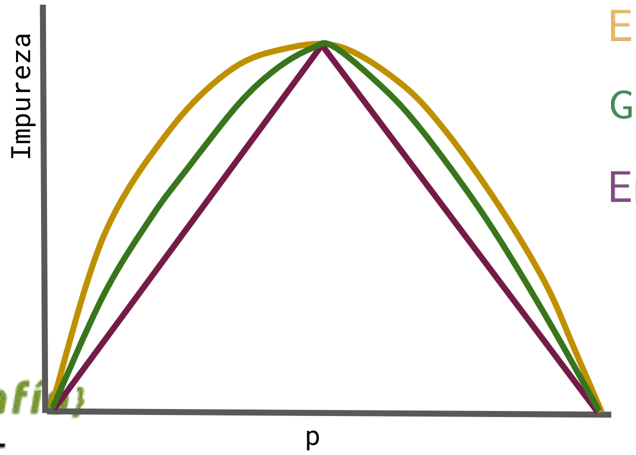


$$\text{Gini} = \sum_{k \neq k'} \hat{p}_{mk}\hat{p}_{mk'} = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$$

$$\text{Error de clasificacion} = \frac{1}{N_m} \sum_{i \in \mathbb{R}_m} I(y_i \neq k(m)) = 1 - \hat{p}_{mk(m)}$$

Impurezas en clasificación

- En un árbol de regresión, el problema se reduce a encontrar una combinación de variable y divisor que reduzca el error cuadrático promedio.
- Para el caso de la clasificación, la medida de impureza conlleva evaluar la tasa de impureza en el modelo.



$$\text{Entropia (o Desviacion)} = -\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk}$$

$$\text{Gini} = \sum_{k \neq k'} \hat{p}_{mk}\hat{p}_{mk'} = \sum_{k=1}^{K} \hat{p}_{mk}(1-\hat{p}_{mk})$$

$$\text{Error de clasificacion} = \frac{1}{N_m}\sum_{i \in \mathbb{R}_m} I(y_i \neq k(m)) = 1 - \hat{p}_{mk(m)}$$

{desafío} latam_

Academia de talentos digitales

www.desafiolatam.com