

A young girl with brown hair, wearing a white headband and white boxing gloves with black accents, is in a boxing stance with her arms raised. She is wearing a dark tank top with a large white letter 'A' on it. The background is a solid light blue.

Algoritmos basados en árboles.

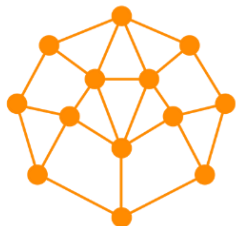
by Saturdays.AI

Saturdays.AI LATAM
Machine Learning



Saturdays.AI
LATAM

Semana 4



Saturdays.AI
LATAM

Agenda

Curso

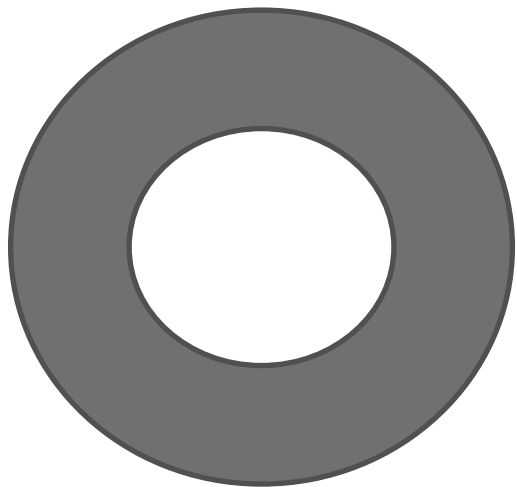
Actividad

Referencias

Curso

Bueno... y qué es un árbol de decisión

Objetivo: Entender el concepto detrás de los algoritmos basados en árboles.



Contenido

- Árboles de decisión
- Criterios de Impuridad
- Ventajas y Desventajas

Curso

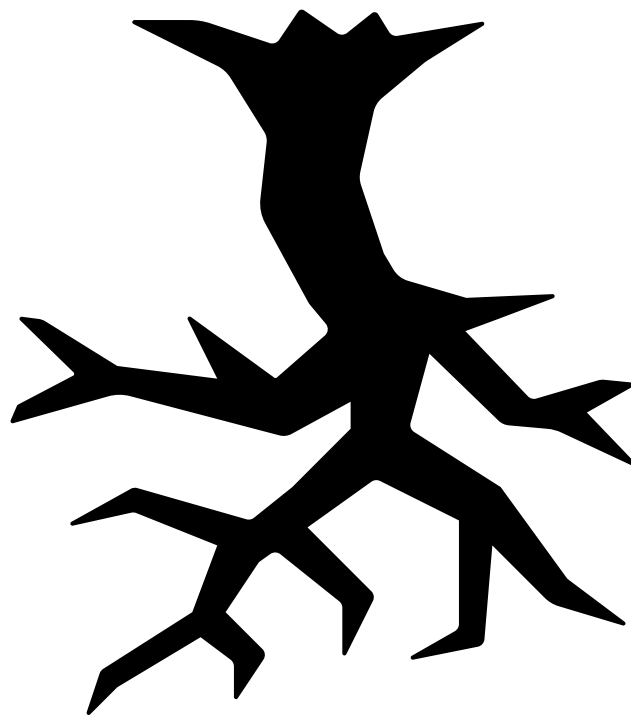
Árboles de decisión:

Modelo no-paramétrico:

- clasificación
- regresión.

Funciona con particiones recursivas de las instancias en el espacio de los datos.

El objetivo de un árbol de decisión es **predecir** el valor a través de un conjunto de **reglas simples** de decisión en su mayoría de casos binaria de las variables otorgadas.



Curso

Árboles de decisión:

¿Crear un árbol?

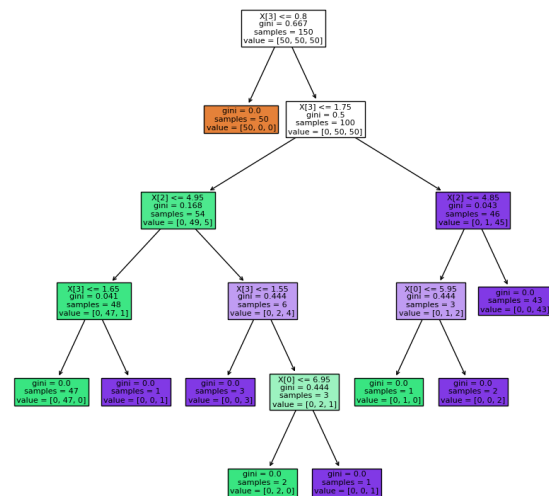
1) Elegir la variable con mayor calidad de la división a través de

-Ganancia de información.

-Impureza Gini

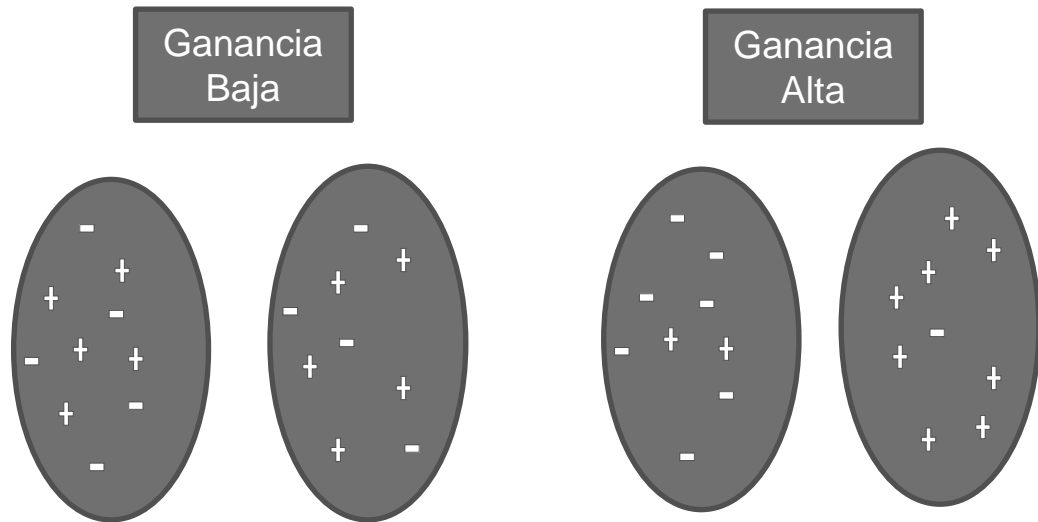
2) Para la siguiente rama, hacer esto recursivamente.

En cada nodos, la variable con mayor información será la que es elegida.



Curso

¡Criterios de impuridad!

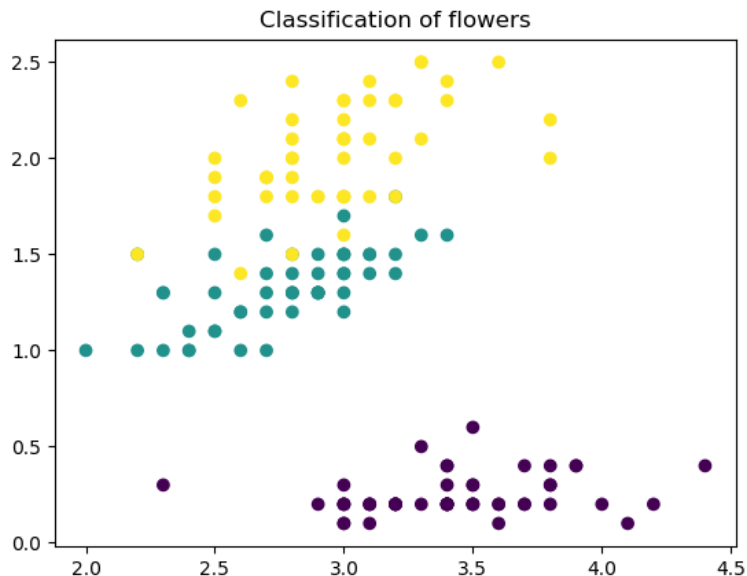


¡**Gini**! Medida de la probabilidad por cada división.

¡**Entropía**! Medida de la aleatoriedad de la información.

ACTIVIDAD

Un poco de python...

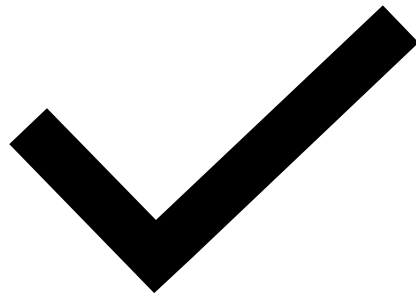


```
|--- feature_3 <= 0.80
|--- class: 0
|--- feature_3 > 0.80
|--- feature_3 <= 1.75
|   |--- feature_2 <= 4.95
|   |   |--- feature_3 <= 1.65
|   |   |--- class: 1
|   |   |--- feature_3 > 1.65
|   |   |--- class: 2
|   |--- feature_2 > 4.95
|   |   |--- feature_3 <= 1.55
|   |   |--- class: 2
|   |   |--- feature_3 > 1.55
|   |   |   |--- feature_0 <= 6.95
|   |   |   |--- class: 1
|   |   |   |--- feature_0 > 6.95
|   |   |   |--- class: 2
|--- feature_3 > 1.75
|   |--- feature_2 <= 4.85
|   |   |--- feature_0 <= 5.95
|   |   |--- class: 1
|   |   |--- feature_0 > 5.95
|   |   |--- class: 2
|   |--- feature_2 > 4.85
|   |--- class: 2
```

Curso

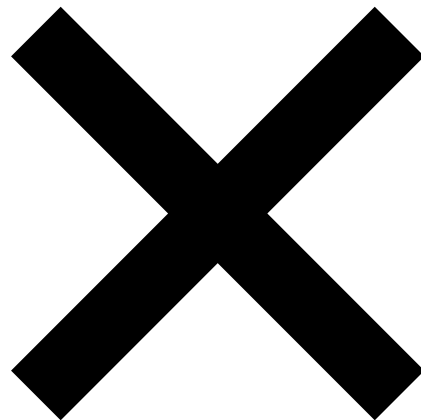
Ventajas

- Abordar relaciones complejas (no lineales).
- Pueden trabajar bien con datos de alta dimensionalidad.
- Requieren mínimo preprocesamiento de datos.
- Robustos ante datos aberrantes
- No tienen sensibilidad ante transformaciones monótonas.
- Métricas para determinar la importancia.



Desventajas

- Tienen un desempeño bajo.
- Poder predictivo está muy limitado.
- Árboles se ven altamente modificados a pequeñas perturbaciones en los datos.
- Pueden tornarse excesivamente complejos.
- Tienden a sobre-ajustar (overfitting).



Curso

Pero... ¿existe una solución alterna?

¡Métodos de Ensamble!

Referencias

- Singh, N. (2019) Decision Tree Algorithm – Explained. Recuperado el 14 de septiembre del 2020 de: <https://towardsdatascience.com/decision-tree-algorithm-explained-83beb6e78ef4>
- Scikit Learn (2020) Decision Trees. Recuperado el 14 de septiembre del 2020 de: <https://scikit-learn.org/stable/modules/tree.html>
- Gupta, P. (2019) Decision Trees in Machine Learning. Recuperado el 14 de Agosto del 2020 de: <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>
- Rokach, L. (n.d.) Decision Trees. Department of Industrial Engineering. Tel-Aviv University. Recupeado el 14 de septiembre del 2020 de: <https://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf>
- Ertel, W. (2011) Introduction to Artificial Intelligence. Springer UTICS Undergraduate Topics in computer science.

Contacto



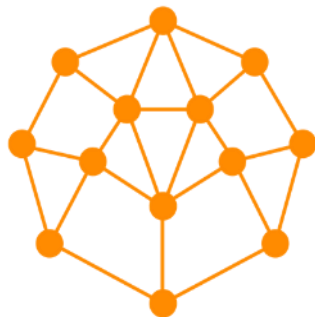
Daniel Hernández Mota:

Daniel@saturdays.ai

www.linkedin.com/in/dhdzmota/



Saturdays.AI
LATAM



Saturdays.AI
LATAM

1ra. Edición Remota

