


Ensamblas de Árboles

Inteligencia de Datos



Agenda


1. Bosques Aleatorios
2. ¿Cuándo usar Árboles de Decisión y Ensamblados?



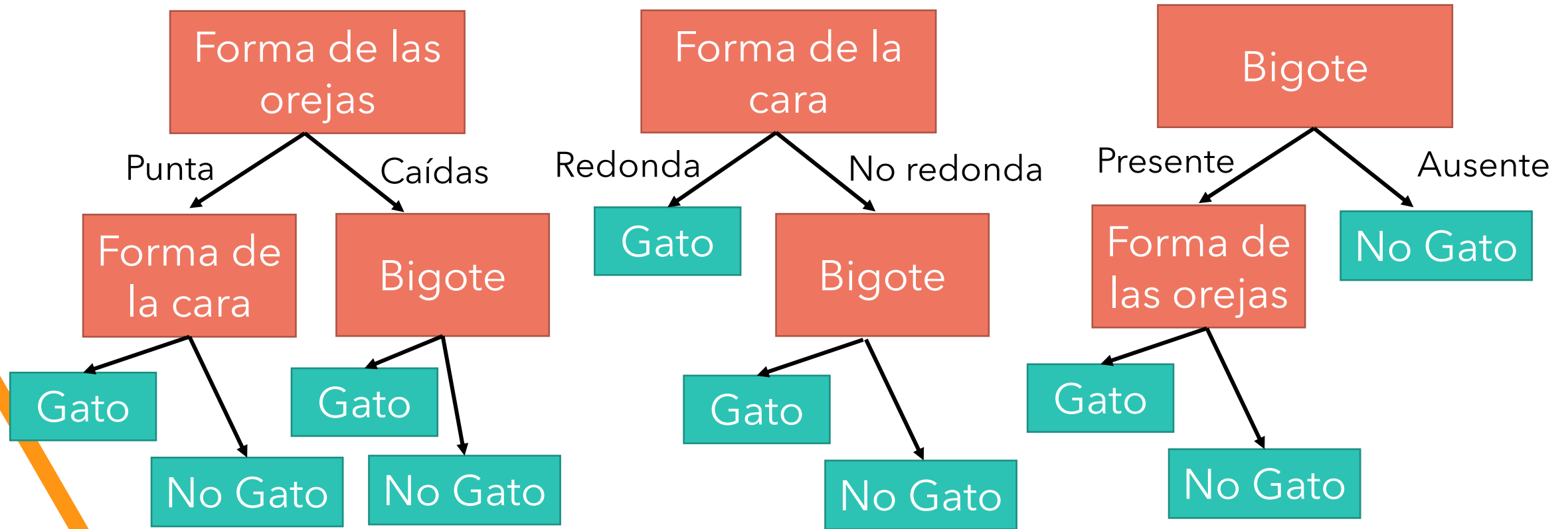
Bosques Aleatorios



Árboles de Decisión

- Los Árboles de Decisión (AD) tienen un enfoque **basado en reglas**.
 - En general, son **muy sensibles a cambios en la información**.
 - ¿Por qué creen que sucede esto?
 - Las particiones pueden cambiar por completo y generar AD diferentes.
 - ¿Cómo podrían atacar este problema con lo que hemos visto hasta el momento?
- 

Ensamble de Árboles - Votación



Algoritmo de Bosques Aleatorios

Dado un conjunto de datos de tamaño \underline{n} .

1. Para $b = 1$ hasta B :
 - a) Obtener un subconjunto mediante muestreo con reemplazo de tamaño \underline{n} .
 - b) Entrenar un AD en el subconjunto creado.
2. Realizar la votación (voto duro o suave).



B controla el tamaño del bosque. Entre más, mejor. Pero, llega un punto en que ya no aumenta y desciende su rendimiento.

Es posible que muchos AD tengan las mismas características para partir un nodo. ¿Cómo se evita?

Algoritmo de Bosques Aleatorios

Aleatorizar la selección de la característica

- En cada nodo de decisión, para elegir la características para realizar la partición...
- si m características están disponibles, al azar determinar un subconjunto $k < m$...
- para que elija de entre ellas (se calcula la mejor partición con el criterio de entropía).
- Para m grande, $k = \sqrt{m}$



¿Cuándo usar Árboles de Decisión y Ensamblés?

¿Cuándo usar Árboles de Decisión y Ensamblados?

Árboles y Ensamblados

- Funcionan **bien** cuando se trabaja con **información tabulada** (estructurada).
- **No son buenos con información sin estructura** como imágenes, texto o audio.
- Su entrenamiento es **rápido**.
- Dependiendo de la profundidad, **pueden ser interpretables**.
- Es mejor usar ensamblados, en general.

Redes Neuronales

- Funcionan **bien con todo tipo de información** (estructurada y no estructurada).
- **Tardan** más en entrenar.
- Existe **transfer learning**.
- Su entrenamiento permite **pegar** varios sistemas de redes neuronales.
- **Muchas opciones** para trabajar con distintos tipos de información.
- **Nada interpretables**.



¿Gracias?

Luis Zúñiga

luis.zuniga@correo.uia.mx