
RESULTADO DE APRENDIZAJE

RdA de la asignatura:

- **RdA 2:** Aplicar modelos de aprendizaje automático supervisado y no supervisado, así como su validación y optimización, en la resolución de problemas tanto reales como simulados.

RdA de la actividad:

- Comprender el funcionamiento de los métodos combinados como bagging, boosting y bosques aleatorios.
- Implementar bosques aleatorios y boosting en un entorno práctico, evaluando resultados.

INTRODUCCIÓN

Pregunta inicial: ¿Por qué combinar varios modelos puede mejorar el rendimiento de un modelo de aprendizaje automático?

DESARROLLO

Actividad 1: Clase magistral sobre métodos combinados

En esta actividad, los estudiantes conocerán las bases teóricas y conceptuales de los métodos combinados en aprendizaje automático: bagging, boosting y bosques aleatorios. Se destacará cómo estos métodos combinan múltiples modelos individuales para mejorar la precisión, reducir la varianza y manejar problemas de sobreajuste. Se usarán ejemplos y discusiones guiadas para consolidar los conceptos.

¿Cómo lo haremos?

- **Exposición teórica:** Presentación de los conceptos clave:
 - **Bagging (Bootstrap Aggregating):** Cómo se generan múltiples modelos a partir de muestras bootstrap para reducir la varianza. Se destacará el uso de árboles de decisión como modelo base.
 - **Boosting:** Explicación del aprendizaje secuencial, donde cada modelo corrige los errores de los anteriores.
 - **Bosques aleatorios:** Introducción a la técnica de combinación de árboles de decisión, resaltando el uso de selección aleatoria de atributos para mejorar la generalización.

Verificación de aprendizaje:

- ¿Qué diferencias fundamentales existen entre bagging y boosting?
- ¿Cómo ayudan los bosques aleatorios a reducir el sobreajuste en comparación con un único árbol de decisión?
- ¿Cuáles son las principales ventajas de combinar múltiples modelos en problemas de aprendizaje automático?

Actividad 2: Bosques Aleatorios

En esta actividad, los estudiantes implementarán un modelo de bosques aleatorios para explorar cómo combina múltiples árboles de decisión para reducir la varianza y mejorar la precisión. Se analizará el impacto de los hiperparámetros, como el número de estimadores y la selección aleatoria de atributos, en el rendimiento del modelo.

¿Cómo lo haremos?

- **Implementación práctica:** Los estudiantes accederán a un cuaderno de Jupyter preparado para esta actividad. El cuaderno incluye un ejemplo de clasificación utilizando un conjunto de datos real y simulado.

Enlace al cuaderno: [17_1-Bosques-Aleatorios.ipynb](#).

Verificación de aprendizaje:

- ¿Qué parámetros afectan el rendimiento de un modelo de bosque aleatorio?

Actividad 3: Boosting y Bagging

Los estudiantes implementarán modelos de boosting y bagging para entender cómo estas técnicas combinan modelos débiles para mejorar el rendimiento. Se compararán las diferencias en el enfoque de aprendizaje (paralelo vs. secuencial) y su impacto en sesgo y varianza.

¿Cómo lo haremos?

- **Implementación práctica:** Los estudiantes trabajarán con un cuaderno de Jupyter para implementar un ejemplo básico de bagging y boosting, ajustando parámetros como el número de estimadores y la tasa de aprendizaje.

Enlace al cuaderno: [17_2-Boosting-Bagging.ipynb](#).

CIERRE

Tarea: Implementar un Bosque aleatorio utilizando un conjunto de datos real y comparar los resultados al variar los criterios de partición, poda y profundidad máxima.

Pregunta de investigación:

1. ¿Cómo influyen los métodos combinados en la detección de anomalías en datasets con ruido?
2. ¿Cómo podrían los métodos combinados adaptarse a datasets dinámicos o en streaming?