

# Teoria-Resuelta-Examen-Enero-202...



**user\_2335920**



**Especialidad: Sistemas de Información**



**4º Grado en Ingeniería Informática**



**Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación  
Universidad de Granada**



[Accede al documento original](#)

antes



**Descarga sin publi  
con 1 coin**



Después

**WUOLAH**



Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato  
→ Planes pro: más coins

perdo  
espacio



Necesito  
concentración

ali ali ooh  
esto con 1 coin me  
lo quito yo...

WUOLAH

(1 pto.) Describe brevemente los modelos básicos de ensemble learning Bagging y Boosting explicando los aspectos relevantes de cada uno de ellos y destacando sus diferencias.

### Bagging (Bootstrap Aggregating)

Bagging es un modelo de *ensemble learning* que combina múltiples clasificadores independientes para reducir la varianza y mejorar la estabilidad y precisión del modelo final.

- **Funcionamiento:**
  - Genera múltiples subconjuntos de datos mediante muestreo con reemplazo (*bootstrap*) del conjunto de datos original.
  - Entrena un modelo base (como árboles de decisión) en cada subconjunto.
  - Combina las predicciones de los modelos mediante votación (clasificación) o promedio (regresión).
- **Características principales:**
  - Reduce la varianza del modelo, mitigando el sobreajuste.
  - Los modelos individuales son entrenados de manera independiente, lo que permite su ejecución en paralelo.
  - Ejemplo clásico: **Random Forest**.

### Boosting

Boosting es un modelo de *ensemble learning* que combina clasificadores débiles secuencialmente, asignando mayor peso a los errores cometidos por los modelos anteriores para mejorar el rendimiento.

- **Funcionamiento:**
  - Entrena un modelo base sobre el conjunto de datos original.
  - Ajusta el modelo siguiente dando mayor peso a las instancias mal clasificadas por el modelo previo.
  - Combina las predicciones de los modelos mediante un esquema ponderado.
- **Características principales:**
  - Reduce el sesgo y mejora la precisión.
  - Los modelos son entrenados de forma secuencial, lo que puede ser computacionalmente más costoso.
  - Ejemplo clásico: **AdaBoost, Gradient Boosting**.

WUOLAH

## Diferencias principales:

Aspecto	Bagging	Boosting
<b>Estrategia</b>	Combina modelos independientes.	Combina modelos secuenciales.
<b>Objetivo</b>	Reduce la varianza (mitiga sobreajuste).	Reduce el sesgo y mejora la precisión.
<b>Entrenamiento</b>	Paralelo (modelos independientes).	Secuencial (modelos interdependientes).
<b>Robustez a ruido</b>	Más robusto frente a datos ruidosos.	Menos robusto al ruido, puede sobreajustar.
<b>Ejemplos populares</b>	Random Forest	AdaBoost, Gradient Boosting.

Ambos enfoques buscan mejorar la precisión del modelo, pero se aplican según el problema y los objetivos específicos del aprendizaje.

(0,5 ptos.) Indica ventajas e inconvenientes de la selección de características de tipo filtro y envolvente

### Selección de características de tipo filtro

La selección de tipo filtro evalúa la relevancia de las características independientemente del modelo predictivo, utilizando métricas estadísticas o heurísticas (e.g., correlación, chi-cuadrado, información mutua).

#### Ventajas:

1. **Independencia del modelo:** Es rápida y no depende de un algoritmo de aprendizaje, lo que la hace eficiente computacionalmente.
2. **Escalabilidad:** Funciona bien con conjuntos de datos grandes, ya que evalúa las características de manera aislada.
3. **Reducción de complejidad:** Permite reducir rápidamente el número de características antes de entrenar modelos más costosos.
4. **Evita sobreajuste:** Al no involucrar al modelo, reduce el riesgo de ajustarse al ruido en los datos.

#### Inconvenientes:

1. **Falta de interacción entre características:** No considera relaciones complejas entre las características ni su impacto conjunto en el modelo.

2. **Menor precisión:** Puede seleccionar características irrelevantes o descartar importantes debido a su evaluación individual.
  3. **Resultados generales:** Su rendimiento puede no estar optimizado para un modelo específico.
- 

## Selección de características de tipo envolvente

La selección de tipo envolvente utiliza un modelo predictivo para evaluar combinaciones de características, seleccionando aquellas que maximizan el rendimiento del modelo.

### Ventajas:

1. **Orientada al modelo:** Garantiza que las características seleccionadas sean óptimas para el modelo utilizado.
2. **Interacción entre características:** Tiene en cuenta las relaciones entre las características, lo que puede mejorar el rendimiento del modelo.
3. **Mayor precisión:** Suele generar subconjuntos más relevantes para el problema específico.

### Inconvenientes:

1. **Costosa computacionalmente:** Requiere entrenar el modelo repetidamente para diferentes subconjuntos, lo que puede ser lento para conjuntos de datos grandes.
2. **Riesgo de sobreajuste:** Puede ajustarse demasiado a los datos de entrenamiento, especialmente con conjuntos pequeños.
3. **Dependencia del modelo:** Los resultados están estrechamente ligados al modelo elegido, lo que puede limitar su generalización.

**(0,5 pts.) Explica brevemente las similitudes, diferencias y qué retos abordan el aprendizaje incremental y la minería de flujo de datos.**

### Similitudes

1. Ambos procesan los datos de manera continua, permitiendo actualizar modelos o descubrir información sin necesidad de reentrenar desde cero.
2. Se centran en trabajar con datos que llegan de forma secuencial o en tiempo real, en lugar de disponer de un conjunto completo desde el inicio.
3. Priorizan la eficiencia computacional, utilizando algoritmos rápidos y que consumen poca memoria.
4. Comparten el reto de adaptarse a cambios en los datos, como los cambios conceptuales (*concept drift*), donde los patrones previos dejan de ser válidos.

Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins?

Plan Turbo: barato

Planes pro: más coins

pierdo  
espacio



Necesito  
concentración

ali ali ooh  
esto con 1 coin me  
lo quito yo...

WUOLAH

## Diferencias

El aprendizaje incremental se enfoca principalmente en la actualización de modelos predictivos, permitiendo que estos se adapten de manera continua a los nuevos datos. Por su parte, la minería de flujo de datos tiene un propósito más amplio: descubrir patrones, tendencias y anomalías en tiempo real, lo que puede incluir tareas como clasificación, agrupamiento o detección de asociaciones.

Mientras que el aprendizaje incremental actualiza un modelo existente, la minería de flujo de datos busca generar conocimiento a partir de un flujo constante de datos, pudiendo abarcar múltiples análisis simultáneos. Además, el aprendizaje incremental generalmente se centra en mantener un modelo eficiente y actualizado, mientras que la minería puede requerir la aplicación de ventanas temporales o métodos especializados para identificar patrones en un contexto más dinámico.

## Retos que abordan

El aprendizaje incremental busca adaptarse a los cambios en los datos (como el *concept drift*) sin olvidar patrones aprendidos previamente, manteniendo la eficiencia en términos de memoria y tiempo. También enfrenta el desafío de evitar el sobreajuste al aprender de nuevos datos.

La minería de flujo de datos, por su parte, aborda el problema de manejar grandes volúmenes de datos que llegan a alta velocidad, extrayendo información útil en tiempo real. Además, debe lidiar con cambios conceptuales, evolución de nuevos patrones (*concept evolution*), y la detección de anomalías o datos ruidosos sin comprometer la calidad de los resultados.

## Conclusión

Ambos enfoques comparten retos comunes relacionados con la naturaleza dinámica de los datos en tiempo real, pero mientras el aprendizaje incremental está diseñado para mantener y actualizar modelos predictivos, la minería de flujo de datos tiene un objetivo más general, centrado en extraer conocimiento útil de los datos de manera continua y eficiente.

WUOLAH