
RESULTADO DE APRENDIZAJE

RdA de la asignatura:

- **RdA 1:** Plantear los conceptos fundamentales del aprendizaje automático, incluyendo los principios básicos, técnicas de preprocesado de datos, métodos de evaluación y ajuste de modelos, destacando su importancia en el análisis y resolución de problemas de datos.

RdA de la clase:

- Diferenciar entre métricas de validación interna y externa en modelos no supervisados.
- Comprender y calcular los conceptos de Cohesión y Separación.
- Interpretar y aplicar el Coeficiente de Silueta para evaluar la calidad de una partición.

INTRODUCCIÓN

Pregunta inicial: En ausencia de etiquetas verdaderas, ¿cómo definimos matemáticamente que un grupo de datos es «bueno» o «compacto»?

DESARROLLO

Actividad 1: Socialización de la Clase Invertida

Revisión de los conceptos trabajados individualmente con ChatGPT sobre la calidad de las particiones.

¿Cómo lo haremos?

- **Dinámica de discusión:** Se seleccionarán estudiantes al azar para definir los siguientes conceptos basándose en su interacción previa con el asistente virtual:
 1. El diámetro de un grupo (cohesión): Discusión sobre la distancia máxima entre puntos del mismo grupo y su relación con la compacidad.
 2. La separación entre grupos: Análisis de la distancia entre centroides o entre los puntos más cercanos de grupos distintos.
 3. Relación Calidad/Métrica: ¿Por qué buscamos minimizar el diámetro y maximizar la separación simultáneamente?

Actividad 2: Métricas de calidad general

Clase magistral breve para formalizar la métrica que combina los conceptos anteriores.

¿Cómo lo haremos?

- **Clase magistral:** Se explicarán las siguientes métricas: Índice de Dunn, Coeficiente de Silueta y el índice de Davies-Bouldin.
- **Implementación en Python:** Los estudiantes accederán a un cuaderno de Jupyter previamente preparado.

Enlace al cuaderno: [05-2-Evaluacion-de-Modelos-Agrupamiento.ipynb](#).

- **Experimentación:**

- Probar las métricas con diferentes números de agrupamientos, ¿cómo varían los valores?

CIERRE

Verificación de aprendizaje:

1. ¿Cómo se calcula la cohesión (inercia) de un clúster y el diámetro de un clúster?
2. ¿Cómo se calcula la separación entre dos clústeres?
3. Si obtengo un Coeficiente de Silueta global de 0.85, ¿qué indica sobre la estructura de mis grupos?
4. ¿Qué sucede con el valor de la cohesión (inercia) a medida que aumentamos el número de clústeres K hasta que $K = N$ (número de datos)?

Preguntas tipo entrevista:

1. Tienes un conjunto de datos con alta dimensionalidad. ¿El Coeficiente de Silueta es confiable o deberías usar otra métrica?
2. ¿Cuál es el coeficiente de silueta de un conjunto en el que solo existe un único clúster?

Pregunta de investigación:

1. ¿Qué son las métricas de validación externa (ej. Índice de Rand, Información Mutua Ajustada) y cuándo se pueden utilizar?

Para la próxima clase: Revisar estos videos:

- [Análisis de componentes principales \(PCA\)](#)
- [SVD Visualized, Singular Value Decomposition explained](#)