



Módulo 3

Sesión N° 1



ACTIVIDAD:



Manipulación de Estructuras NumPy

Objetivo: Manipular arreglos multidimensionales en NumPy aplicando selección condicional, operaciones agregadas, álgebra lineal básica, limpieza de datos, análisis de correlación y reducción de dimensionalidad.



Contexto

Como analista de datos para un laboratorio de ingeniería, recibes una base de datos de 30 sensores industriales, cada uno midiendo 96 valores horarios de tres variables distintas: temperatura (°C), presión (kPa) y humedad (%). Los datos presentan valores atípicos, valores perdidos y patrones que requieren pre-procesamiento y análisis exploratorio profundo.

Debes realizar, con NumPy, un procesamiento avanzado de estos datos en una matriz tridimensional y responder a retos de análisis y transformación que combinan álgebra lineal, filtrado condicional, operaciones vectorizadas y resumen estadístico.



Requerimientos técnicos

- Python 3.x y NumPy (puedes usar Jupyter, Colab o IDE de preferencia).
- El código debe estar correctamente comentado, bien estructurado y todas las respuestas deben tener justificación breve.
- Entrega: archivo .py, .ipynb o notebook de Colab con código, salidas y respuestas argumentadas.

Entregable:

- Notebook o script comentado, salidas, respuestas numéricas y justificaciones.
- Modalidad de trabajo: grupal.
- Tiempo estimado de desarrollo: 75 minutos.





Requerimientos:

1. Genera un arreglo 3D llamado D de tamaño (30, 96, 3), con datos aleatorios de sensores:
 - Temperatura: entre 5°C y 110°C
 - Presión: entre 80 y 140 kPa
 - Humedad: entre 15% y 85%

(Usa `np.random.uniform` para cada canal y concatena por la tercera dimensión)
2. Introduce valores perdidos y atípicos:
 - Escoge aleatoriamente 30 valores del arreglo total y asígnales `np.nan` (simulando sensores defectuosos).
 - Introduce valores extremos (3 valores mayores al rango superior y 3 menores al rango inferior en cada canal).
3. Limpieza inicial:
 - Identifica y cuenta el número de valores perdidos por variable.
 - Sustituye los valores perdidos por la media de la variable correspondiente (por sensor).
4. Filtrado y análisis condicional:
 - Selecciona todos los sensores cuya temperatura máxima registrada supere los 100°C y cuya humedad mínima sea inferior a 20%.
 - ¿Cuántos sensores cumplen ambas condiciones?
5. Normalización y reducción:
 - Normaliza los datos de presión de todos los sensores a rango [0, 1] por sensor.
 - Calcula la media, mediana y desviación estándar de cada variable para estos sensores filtrados.
6. Correlación y análisis avanzado:
 - Para los sensores seleccionados, calcula la matriz de correlación entre las tres variables usando todos sus registros.
 - ¿Qué variable está más correlacionada con la temperatura en este subconjunto?
7. Álgebra lineal aplicada:
 - Para el sensor con mayor varianza en temperatura, calcula la matriz de covarianza entre sus tres variables.
 - Realiza una descomposición en valores y vectores propios de la matriz de covarianza, e interpreta el resultado (¿qué variable aporta más a la variabilidad?)

