



Módulo II. Métodos de Machine Learning

CURSO:

**Taller de Machine Learning
para el análisis y visualización
en Power BI**

Laboratorio de Inteligencia Artificial

Tema 03 – Aprendizaje no supervisado

Profesor: Saúl Domínguez Isidro, PhD.
Contacto: saul.dominguez@lania.edu.mx

Objetivo

Entender los principales conceptos
y algoritmos de aprendizaje no
supervisado



Contenido

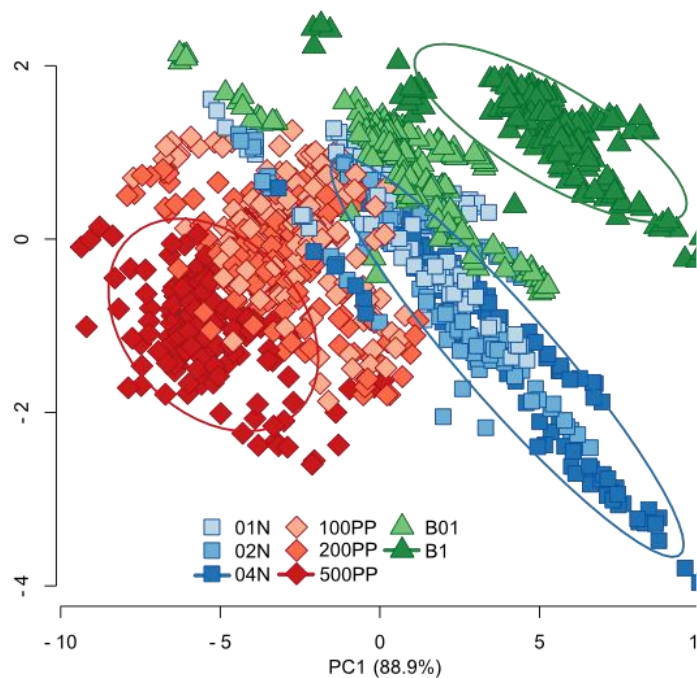
- Introducción
- Transformación
- Reducción de dimensionalidad
- Agrupamiento

Aprendizaje no supervisado

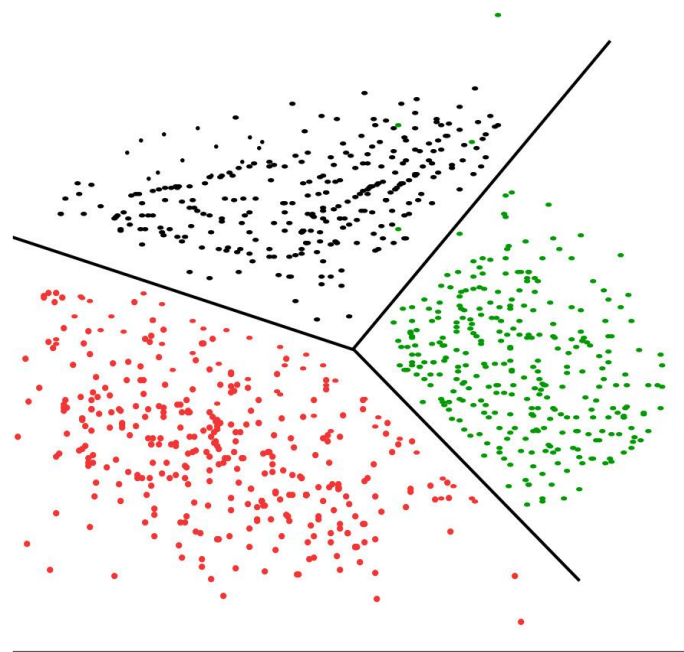
- El aprendizaje no supervisado subsume todo tipo de aprendizaje automático donde no hay resultados conocidos, ninguna retroalimentación para instruir el algoritmo de aprendizaje.
- En el aprendizaje no supervisado, al algoritmo de aprendizaje solo se le muestran los datos de entrada y se le pide que extraiga el conocimiento de un conjunto de datos.

Tipos de aprendizaje no supervisado

Unsupervised transformations

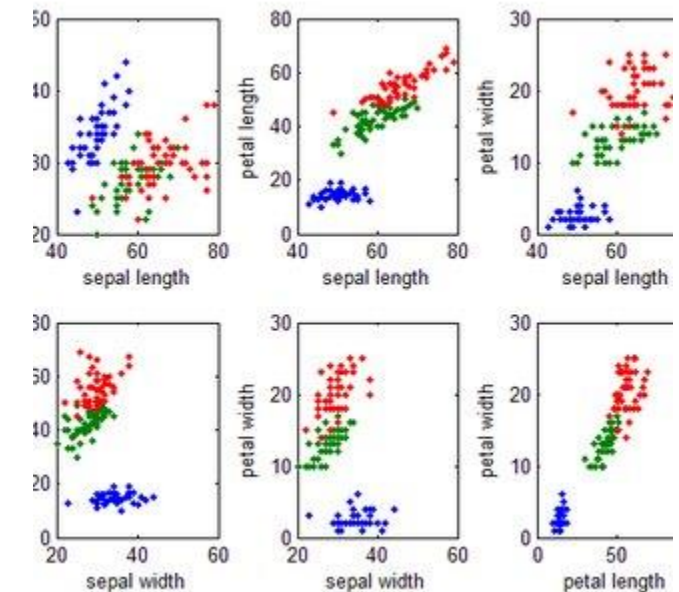


Clustering algorithms



Retos del aprendizaje no supervisado

- Un desafío importante en el aprendizaje no supervisado es evaluar si el algoritmo aprendió algo útil.
- Los algoritmos de aprendizaje no supervisados generalmente se aplican a datos que no contienen información de etiqueta, por lo que no sabemos cuál debería ser el resultado correcto.



Retos del aprendizaje no supervisado

- Como consecuencia, los algoritmos no supervisados se utilizan a menudo en un entorno exploratorio, cuando un científico de datos desea comprender mejor los datos, en lugar de como parte de un sistema automático más grande.
- Otra aplicación común de los algoritmos no supervisados es como paso de preprocesamiento para los algoritmos supervisados.
- Aprender una nueva representación de los datos a veces puede mejorar la precisión de los algoritmos supervisados o puede conducir a una reducción del consumo de memoria y tiempo.

Preprocesamiento y escalado

- El escalamiento de características consiste en transformar los valores de diferentes características numéricas para que se encuentren dentro del rango similar.
- Cuando un conjunto de datos se aplica en algoritmos como optimización del descenso de gradiente o los KNN, el algoritmo intenta y encuentra optimizar pesos o distancias para manejar valores de características que tienen valores más grandes.

¿Porqué utilizar el escalado?

- Resultando en modelos que son de naturaleza subóptima.
- Aquí es donde entra en juego la escala de características. La idea es transformar el valor de las características en un rango similar como otras para que los algoritmos de aprendizaje automático se comporten mejor dando como resultado modelos óptimos.
- El escalado de características no es importante para algoritmos como el Random Forest o árboles de decisión, que son invariantes en el escalado. La escala del valor de las características no afecta el rendimiento del modelo de los modelos entrenados con estos algoritmos.

Normalización

- se refiere al cambio de escala de las características a un rango de [0, 1], que es un caso especial de escala mínima-máxima.

$$x_{norm}^{(i)} = \frac{x^{(i)} - x_{min}}{x_{max} - x_{min}}$$

Estandarización

- Se utiliza para centrar las columnas de características en la media 0 con una desviación estándar 1 para que las columnas de características tengan los mismos parámetros que una distribución normal estándar. A diferencia de la normalización, la estandarización mantiene información útil sobre los valores atípicos y hace que el algoritmo sea menos sensible a ellos en contraste con el escalado mínimo-máximo, que escala los datos a un rango limitado de valores.

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu_x}{\sigma_x}$$

Reducción de dimensionalidad: Uso

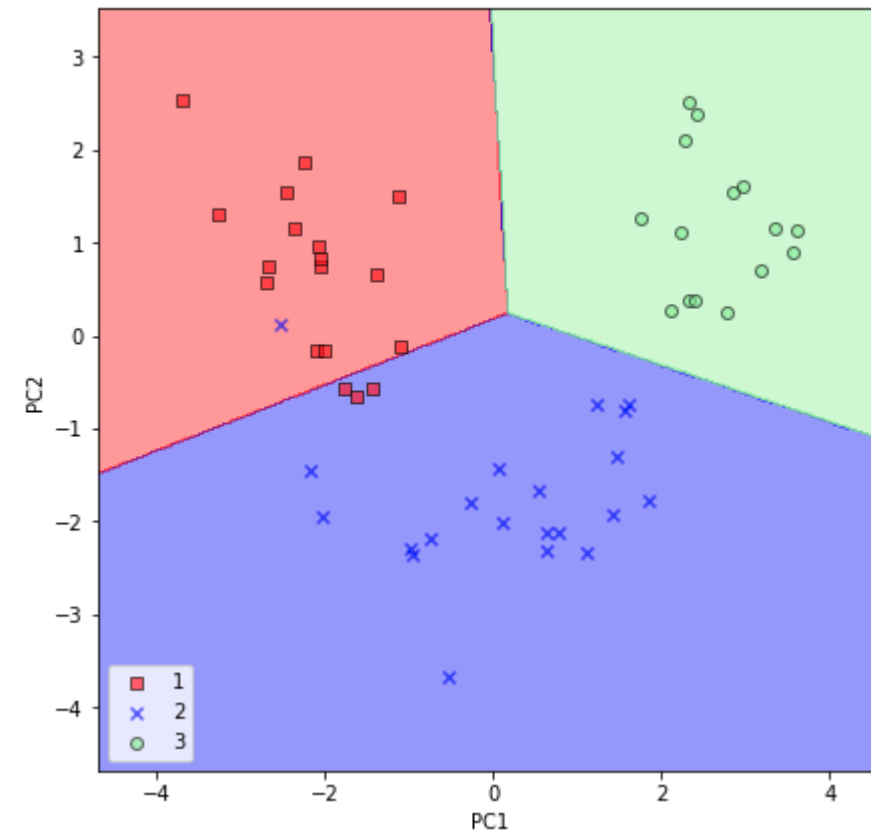
- **Identificar y eliminar las variables irrelevantes.**
- **No siempre el mejor modelo es el que más variables tiene en cuenta.**
- **Se mejora el rendimiento computacional**, lo que se traduce en un ahorro en coste y tiempo.
- **Se reduce la complejidad**, lo que lleva a facilitar la comprensión del modelo y sus resultados.

Análisis de componentes principales (PCA)

- es una técnica de transformación lineal no supervisada que se usa ampliamente en diferentes campos, principalmente para la extracción de características y la reducción de dimensionalidad
- PCA ayuda a identificar patrones en los datos basados en la correlación entre características.

Análisis de componentes principales (PCA)

- Si usamos PCA para la reducción de dimensionalidad, construimos una matriz de transformación $d \times k$ -dimensional W , que nos permite mapear un vector de muestra x en un nuevo subespacio de características k -dimensional que tiene menos dimensiones que el espacio original de características d -dimensionales



K-means

- El **clustering** es una técnica para encontrar y clasificar K grupos de datos (clusters). Así, los elementos que comparten características semejantes estarán juntos en un mismo grupo, separados de los otros grupos con los que no comparten características.
- Para saber si los datos son parecidos o diferentes el algoritmo K-medias utiliza la distancia entre los datos. Las observaciones que se parecen tendrán una menor distancia entre ellas. En general, como medida se utiliza la distancia euclideana aunque también se pueden utilizar otras funciones.

K-means: Funcionamiento

- K-means necesita como dato de entrada el número de grupos en los que vamos a segmentar la población. A partir de este número k de clusters, el algoritmo coloca primero k puntos aleatorios (centroides). Luego asigna a cualquiera de esos puntos todas las muestras con las distancias más pequeñas.
A continuación, el punto se desplaza a la media de las muestras más cercanas. Esto generará una nueva asignación de muestras, ya que algunas muestras están ahora más cerca de otro centroide. Este proceso se repite de forma iterativa y los grupos se van ajustando hasta que la asignación no cambia más moviendo los puntos. Este resultado final representa el ajuste que maximiza la distancia entre los distintos grupos y minimiza la distancia intragrupo.

K-means: Funcionamiento

- Recibe como parámetro k clusters para segmentar la población.
- Coloca k puntos aleatorios (centroides)
- Hasta alcanzar condición de paro
 - Asignar a cada punto todas las muestras con distancias pequeñas
 - Desplazar el punto a la media de las muestras más cercanas
- Fin

El resultado final representa el ajuste que maximiza la distancia entre los distintos grupos y minimiza la distancia intragrupo.

Ejemplo

