

Universidad Nacional Autónoma de Honduras
Facultad de Ingeniería
Departamento de Ingeniería en Sistemas



Inteligencia Artificial
III PAC – 2024

Guía de ejercicios

1. Clasificación de dígitos con PCA y KMeans

Reducir la dimensionalidad de un conjunto de imágenes de dígitos escritos a mano, como el dataset digits de sklearn (df = load_digits), utilizando PCA.

Luego, aplicar KMeans para agrupar los dígitos en clústeres y ver cómo se relacionan con sus etiquetas reales.

El dataset mencionado anteriormente, contiene los pixeles de imágenes 28x28, que representan los dígitos del 0 al 9.

- Con lo anterior usted deberá reducir la dimensionalidad de las imágenes de 784 pixeles a un número mucho menor, 50 componentes principales.
- Observar la variabilidad de los datos explicada por cada componente y elegir el número de componentes que retenga al menos el 90% de la varianza.
- Cuando ya se obtenga el resultado de los PCA principales aplicar algoritmo KMeans a los datos reducidos con PCA para dividirlos en 10 clústeres, ya que sabemos que hay diez dígitos.
- Asigna los dígitos originales a los clústeres generados y evaluar cuántos dígitos coinciden correctamente en el mismo grupo.
- Utilizar métricas como la exactitud o el puntaje de silueta para evaluar la calidad de los clústeres.
- Visualizar los dígitos proyectados en 2D utilizando solo los primeros dos componentes principales y colorea los puntos por clústeres. Esto te permitirá observar visualmente la separación de cada grupo.

2. Agrupamiento de datos de las flores con el data set Iris.

Utilizar KMeans para agrupar el conjunto de datos **Iris**, que contiene información sobre tres especies de flores (Iris-setosa, Iris-versicolor, e Iris-virginica). Los grupos se formarán en función de las características de cada flor.

- Cargue el dataset Iris desde sklearn.datasets.
- Aplica el algoritmo KMeans para agrupar las muestras en tres clústeres (correspondientes a las tres especies).
- Comparar los clústeres generados con las etiquetas reales de las especies de flores para evaluar la precisión del modelo.
- Visualizar los resultados en un gráfico de dispersión de dos o tres características, coloreando los puntos según el clúster asignado.

3. Segmentación de Clientes para un Análisis de Marketing

Utilizar KMeans para realizar una **segmentación de clientes** en función de sus patrones de compra y otros datos demográficos. La segmentación ayudará a identificar grupos de clientes para campañas de marketing personalizadas.

- Cree un data set con datos simulados que incluya ingresos anuales, gasto en productos, edad, género.
- Normaliza las características para que los datos estén en el mismo rango, lo cual mejora el rendimiento del algoritmo.
- Usa KMeans para dividir los clientes en cinco clústeres, representando diferentes segmentos de clientes.
- Visualiza los segmentos de clientes en función de dos o tres características y analiza los clústeres.

4. Detección de Anomalías en Transacciones Financieras

Utilizar KMeans para detectar posibles **anomalías** en transacciones financieras de una empresa. Las transacciones inusuales pueden corresponder a fraudes o errores, y la agrupación ayudará a identificar patrones anómalos.

Cree un dataset a partir de estos datos:

{'Monto': [200, 220, 250, 3000, 500, 550, 8000, 600, 2500, 7000], 'Frecuencia': [2, 3, 5, 20, 7, 4, 50, 6, 30, 45]}

- Usa KMeans para identificar clústeres de transacciones típicas.
- Calcula la distancia de cada transacción al centroide de su clúster y marca las transacciones que están fuera de cierto umbral como posibles anomalías.
- Visualiza las transacciones normales y las anómalas en un gráfico.