

Tarea 4

Problema

Considere el conjunto de datos *Countries of the world*¹, que presenta indicadores, la mayoría continuos, de distintos países del mundo.

Se pide:

a) Implemente el algoritmo k-means visto en el teórico. Seleccione algunos de los atributos continuos que le parezcan relevantes y agrupe los países en diferentes números de clusters. Utilice el método del codo para obtener el número “óptimo” de clusters. Comente sus resultados.

b) Utilice PCA para visualizar los clusters construidos. Identifique en la visualización la región de cada país. Analice en forma cualitativa la relación entre las regiones y los clusters construidos. ¿Qué países comparten cluster con Uruguay? Comente brevemente.

Se puede utilizar numpy, scipy, pandas y scikit-learn para cualquier tarea (incluida PCA) que no esté relacionada con el algoritmo de clustering que se pide implementar.

Entregables

- Informe con las pruebas realizadas y los resultados obtenidos.
- El informe a entregar debe ser un Jupyter Notebook.
- Código escrito para resolver el problema.

Fecha límite de entrega

Lunes 1ero. de noviembre (inclusive)

¹<https://www.kaggle.com/fernandol/countries-of-the-world>