Práctico 4
Aprendizaje no supervisado

Ejercicio 1
Considere el siguiente conjunto de datos en $\mathbb{R}^2$

```
(1.89  56.95),(2.02  53.64),(1.82  57.27),(1.59  62.81),(1.86  58.28),
(1.78  61.08),(1.63  64.93),(1.47  64.42),(1.79  58.81),(1.84  58.94),
(1.79  59.71),(1.68  61.33),(1.76  62.17),(1.77  59.55),(1.79  59.39),
(1.77 56.35),(1.83 59.17),(1.79 60.7 ),(1.73 61.03),(1.71 59.98)
```

Siguiendo el procedimiento para el cálculo de PCA visto en el teórico:

a) Calcule la varianza de cada una de las dimensiones
b) Grafique el conjunto original
c) Grafique el conjunto luego de restar la media a cada una de las instancias. Verifique que la nueva media es 0.
d) Calcule la matriz de covarianza del nuevo conjunto de datos
e) Obtenga los valores y vectores propios de la matriz de covarianza
f) Manteniendo las dimensiones, ajuste el conjunto de datos para que la base sean ahora las direcciones ordenadas por su varianza. Grafique el resultado
e) Repita el procedimiento, pero reduciendo la dimensión a 1. Grafique.
f) Invierta el procedimiento, obteniendo los datos originales, pero solo con el componente principal. Compare con el dataset original. ¿Cuál es la diferencia?

Ejercicio 2 (Laboratorio 2019)
a) Utilice PCA para describir el corpus *Aquienvoto.uy*[1],
b) Responda la siguiente pregunta: ¿puede afirmarse que los partidos políticos agrupan votantes con respuestas similares?

Ejercicio 3
Dé los centroides resultantes de aplicar el algoritmo K-Means con k=3, dados los puntos siguientes:

(1,1), (1.2, 1.2), (1.1, 1), (1.3, 0.9), (3.2, 2.9), (3,3.1), (3,3), (3,1), (2.9, 1.1), (2.8,0.9)

y asumiendo una inicialización de centroides en (1,1), (3,1) y (2.8,0.9).

Ejercicio 4
Aplique el algoritmo k-means al corpus Aquienvoto.uy (solamente teniendo en cuenta las respuestas a las preguntas) para generar n clusters, con n = 2,3,5,10. Analice los resultados obtenidos. ¿Cuál sería el mejor número de clusters? Justifique.

Ejercicio 5
Calcule el Rand Index y el Advanced Rand Index para los dos agrupamientos siguientes (donde cada número indica a qué cluster pertenece el elemento correspondiente):

(1 2 3 3 2 1 1 3 3 1 2 2)
(3 2 3 2 2 1 1 2 3 1 3 1)

1 https://github.com/johnblanco/predictor_electoral