

Fernando José Garavito Ovando 18071

### Instrucciones:

- Deben unirse a uno de los grupos de Canvas de nombre "Laboratorio 3 #", donde N es un número entre 1 y 8. Los grupos pueden ser de 3 personas..
- Esta actividad debe realizarse en grupos.
- Sólo es necesario que una persona del grupo suba el trabajo a Canvas.
- No se permitirá ni se aceptará cualquier indicio de copia. De presentarse, se procederá según el reglamento correspondiente.
- Deberá crear un video de no más de 10 minutos explicando y presentando su solución.
  - Recuerde usar los conceptos vistos en clase.
- Tendrán que responder un cuestionario a modo de reflexión individual tras finalizar el laboratorio.

### Task 1 - K-Means

Usando el dataset dado, implemente un modelo de K-Means. Para esto tenga en cuenta lo siguiente

- Seleccione el K que mejor se adapte al dataset. Justifique su elección
- Muestre al menos una gráfica que muestre cómo se dividen los clusters seleccionados
- Se recomienda el uso de librerías, pero si gusta puede implementar el código por su cuenta

### Task 2 - Mixture-Models

Usando el dataset dado, implemente un modelo de Mixture-Models. Para esto tenga en cuenta lo siguiente

- Seleccione el número de clusters que mejor se adapte al dataset. Justifique su elección
- Muestre al menos una gráfica que muestre cómo se dividen los clusters seleccionados
- Se recomienda el uso de librerías, pero si gusta puede implementar el código por su cuenta

### Task 3 - Preguntas teóricas

En un archivo PDF responda las siguientes preguntas

- ¿Qué diferencias resultan de los clusters dados en K-Means y Mixture-Models para el dataset utilizado?
- ¿Por qué cree que se dan estas diferencias?
- ¿Ha variado la selección del número de clusters entre K-Means y Mixture-models? ¿Por qué?
- ¿En qué casos usaría K-means? ¿Por qué?
- ¿En qué casos usaría Mixture-Models? ¿Por qué?

### Respuestas

1.
  - K Mean detecta formas esféricas mientras que GMM puede con elípticas.
  - GMM calcula la distancia ponderada y K Means calcula la distancia euclidiana convencional.
  - Uno es minimizar mientras el otro es agrupamiento.
2. Ambas cuentan con su respectivo calculo y existe la variante del denominador para la varianza.
3. Me parece que los cluster de KMeans son mucho mas confiables que con GMM.
4. KMeans lo usaría en grandes cantidades de datos como ventas, inventarios, actividades para ciudadanos, etc. Esto debido a que es mucho mas sencillo detectar los datos y la cantidad excesiva que contienen.
5. GMM lo usaría para probabilidades, matrices, etc. Esto debido a su distribución al calcular.

1. Link al repositorio de los integrantes del grupo.
  - a. Deberán subir el código también a Canvas por temas de Acreditación
2. Link al video del funcionamiento su implementación

## Evaluación

1. [30 pts.] Tasks 1 y 2
  - a. [20 pts.] El código funciona correctamente cumpliendo con el objetivo de cada task
  - b. [5 pts.] Aplicación clara y correcta de los conceptos teóricos y calidad del código
  - c. [5 pts.] Mención adecuada en el video de presentación
2. [20 pts.] Task 3
  - a. [4 pts.] Respuesta a cada pregunta
3. [20 pts.] Reflexión Individual