

Inteligencia Artificial - Laboratorio 3 -

Semestre I - 2022

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

o Video de task 1: https://youtu.be/BQX3GVXnD_U

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?

Las formas de los clusters pueden ser circulares o elípticas en el GMM mientras que k-means asume que las siluetas son circulares.

- ¿Por qué cree que se dan estas diferencias?

Por varias razones, los clusters no necesariamente contienen la misma cantidad de datos, k-means ignora esto mientras que GMM no, también porque en GMM se calcula la varianza y se utiliza para entender y graficar el cluster que mejor se ajusta a los datos.

- ¿Ha variado la selección del número de clusters entre K-Means y Mixture-models? ¿Por qué?

En nuestro dataset no, ambos análisis de codos indicaron que el número óptimo era 3 clusters pero eso no es necesariamente aplicable a todos los datasets. Datasets más dispersos podrían ser mejor ajustados con un

mayor numero de clusters en GMM o con formas elípticas.

- ¿En qué casos usaría K-means? ¿Por qué?

K means se usaría al momento de querer rapidez y sencillez, ya que es un algoritmo mucho mas sencillo, y con resultados óptimos, no siempre dará los mejores resultados, sin embargo, son más fáciles de interpretar.

Y de ser el caso que los clusters no tienen se puede cambiar el acercamiento al problema con GMM.

- ¿En qué casos usaría Mixture-Models? ¿Por qué?

Al graficar los datos (o al menos una muestra representativa) si se puede notar que los datos están dispersos en varias dimensiones o que no tienen formas circulares bloques de ellos, sería una buena estrategia utilizar GMM porque este algoritmo maximizara la probabilidad de que los datos pertenezcan a un cluster incluso si no tienen formas que se puedan delimitar con un círculo.

Entregas en Canvas

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