

Universidad del Valle de Guatemala  
Facultad de Ingeniería  
Departamento de Ciencias de la Computación



## Laboratorio III - IA

Daniela Villamar 19086  
Mirka Monzón 18139  
Alexa Bravo 18831

# Preguntas Teóricas

- **¿Qué diferencias resultan de los clusters dados en K-Means y Mixture-Models para el dataset utilizado?**

En el caso de K-Means este generalmente utiliza un enfoque determinístico y asigna cada punto de dato a un grupo único, al contrario de Mixture-Models el cual es probabilístico y considera cada punto para cada grupo.

K-Means utiliza 2 parámetros, el número de clusters y la locación de los centroides, en el caso de Mixture-Models, se utiliza el número de clusters, la media y la covarianza de los clusters.

La última diferencia entre los clusters es que K-Means requiere una estandarización acorde al centroide si así requieren las dimensiones del cálculo, y en el caso de Mixture-Models automáticamente toma en cuenta el problema del cálculo y el uso de la matriz de covarianza.

- **¿Por qué cree que se dan estas diferencias?**

Creemos que estas diferencias se dan por como están diseñados los modelos de cluster, en el caso de K-Means este se basa en el número de clusters y los centroides, colocando cada datos donde corresponda, hasta terminar, ahora en el caso de Mixture-Models se basa en la media y la covarianza para determinar la posición y la forma de cada cluster, por lo que se entiende cómo es que las diferencias antes dichas ocurren con el mismo dataset.

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

Si ha variado, en el caso de K-Means se utilizó 3 clusters y para Mixture-Models se utilizó 2. Aquí varía porque en el momento de realizar la gráfica de codo, se utiliza el método correspondiente, si es para K-Means se usa el mismo método, al igual que para Mixture-Models este se utiliza para la misma gráfica.

En conclusión al momento de tener dos métodos distintos, la gráfica de codo reacciona de acuerdo a como graficamos después los datos.

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

La agrupación en clústeres de K-means es un algoritmo de aprendizaje automático no supervisado muy potente. Se utiliza para resolver muchos problemas complejos de aprendizaje automático no supervisados. El agrupamiento k-means intenta agrupar tipos similares de elementos en forma de grupos. Encuentra la similitud entre los elementos y los agrupa en grupos. El algoritmo de agrupamiento K-means funciona en tres pasos. Entre sus ventajas encontramos que:

1. Es muy simple de implementar.
2. Es escalable a un gran conjunto de datos y también más rápido a grandes conjuntos de datos.
3. Adapta los nuevos ejemplos con mucha frecuencia.
4. Generalización de clusters para diferentes formas y tamaños.

K-means preferiblemente se puede aplicar a datos que tienen un número menor de dimensiones, son numéricos y son continuos. Por ejemplo, un escenario en el que queremos hacer grupos de cosas similares a partir de una colección de cosas distribuidas al azar, en este caso, k-means es muy adecuado para este tipo de escenarios.

### • **¿En qué casos usaría Mixture-Models? ¿ Por qué?**

Mixture Models es un método de agrupación que utiliza una distribución de probabilidad. La agrupación en clústeres de K-means también es un método de agrupación, pero utiliza la distancia euclidiana para calcular la diferencia entre los puntos de datos, ya que los datos más cercanos se pueden segregar en un clúster, esta es una gran diferencia entre K-mean y Mixture Models. Este método se puede utilizar en el procesamiento de señales para abstraer varias características útiles de los datos de la señal, tiene la capacidad de proporcionar buenos resultados en el reconocimiento de voz. En lugar del procesamiento de señales, también es útil en el procesamiento de imágenes y la visión por computadora, donde se puede usar en la segmentación de imágenes y la detección de anomalías.