|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tipo de actividad | Quiz/Cuestionario | | | |
| Título de la actividad (como se verá en plataforma)  *Aquí se pone el título que el estudiante verá en la plataforma (vista lateral y vista por semanas)* | Generalidades de K-Medias | | | |
| Objetivo(s) de la semana a la que apunta esta actividad  *Estos deben coincidir con la información de la tabla que está más adelante, en la que se verifican preguntas frente a objetivos.* |  | | | |
| ¿Es sumativa?  *Seleccionar con X* | Sí [ ] | No | | |
| Número de intentos máximos (Si es formativo anote NA) |  | | | |
| Tiempo de la actividad  *Horas:Minutos* |  | | | |
| **Tipo de comentario**  *Se refiere a la retroalimentación de la pregunta y de la actividad en general. (seleccionar una opción con [X])* | Completo [ ] | | Parcial [ ] | Limitado [ ] |
| *Completo: el estudiante podrá ver la puntuación total y el porcentaje de preguntas correctas, puntaje por pregunta y estado de cada una (correcta/incorrecta), retroalimentación de las preguntas y la/s rúbricas utilizadas para la calificación.* | | *Parcial: El estudiante dejará de ver la retroalimentación de las preguntas.* | *Limitado: el estudiante únicamente podrá ver la puntuación total y el porcentaje de preguntas correctas*. |
| **Plantilla de cuestionario**  **Opciones múltiples**   1. **Múltiples respuestas correctas** 2. **Única respuesta correcta**   **Respuesta libre**   1. **Coincidencia de texto** 2. **Expresión matemática** 3. **Numérico** 4. **Expresión regular** 5. **Expresión de código**   **Reflexión**   1. **Opción múltiple de reflexión** 2. **Opción única de reflexión** 3. **Respuesta de texto de reflexión**   **Personalizado**   1. **Pregunta de complemento** | *Más adelante en este documento encuentra el formato que debe seguir para preguntas de opción múltiple con única o múltiple respuesta. Si tenemos otras preguntas, podemos revisar para indicar el formato en que debe presentarse.*  *Las preguntas que se llaman de Reflexión dan el punto al estudiante independiente de su respuesta.*  *Aspectos importantes para tener presentes en la construcción de preguntas y opciones de respuesta:*  *1. Cada opción de respuesta debe tener retroalimentación.*  *2. Todas las opciones de respuesta deberían ser lógicas y acordes con el tema estudiado, la idea con estas es propiciar cierta discriminación que permita verificar las comprensiones de los estudiantes.*  *3. La retroalimentación no debe ofrecer la respuesta correcta, sino centrarse en la razón de la equivocación. Esto es más sencillo de lograr cuando se tiene en cuenta el punto anterior. Es difícil crear una retroalimentación formativa para una opción “descabellada” o “ilógica”.*  *4. No se sugiere opciones como ninguna de las anteriores o todas las anteriores.*  *5. Por defecto, está configurado para que las opciones de respuesta se muestren en diferente orden cada vez. Si se necesita mostrar en algún orden específico debemos indicarlo.*  *6. Por defecto, todas las preguntas se ponderan con el mismo valor. Si se quiere algo diferente, debemos indicarlo.*  *7. Por accesibilidad, cuidar que si una pregunta tiene una imagen, el texto explique lo necesario para que quien no accede a la imagen pueda responder la pregunta.* | | | |

**Preguntas con el formato de la plantilla**

**Question 1**

¿Por qué el algoritmo de K-Medias es determinístico?

A: El algoritmo de K-Medias no es determinístico porque siempre que se corre puede llegar a soluciones diferentes.

Feedback: La solución de K-Medias depende de los centroides iniciales.

\*B: El algoritmo de K-Medias depende de los centroides iniciales por lo que la solución a la que se llega puede cambiar en la medida que se escojan otros centroides.

Feedback: Si se tienen los centroides fijos, siempre se llega a la misma solución.

C: El algoritmo de K-Medias depende del número de segmentos K que se escojan inicialmente, por eso se dice que está determinado por K.

Feedback: Por determinístico se entiende que, dados los parámetros iniciales, el algoritmo siempre converge a una misma solución.

D: El algoritmo de K-Medias es determinístico porque para diferentes conjuntos de datos se pueden tener los mismos segmentos.

Feedback: Para cada conjunto de datos, K-Medias encontrará segmentos diferentes.

**Question 2**

¿Por qué el algoritmo de K-Medias no puede recibir como input una matriz de distancias entre las observaciones?

\*A: Porque para calcular los centroides necesita calcular el valor promedio del segmento.

Feedback: Correcto, dado que K-Medias necesita poder encontrar las medias de diferentes subconjuntos de las observaciones que desea agrupar, no es posible usar una matriz de disimilitud como input de este algoritmo. Sin embargo, una variación de K-Medias llamada K-Medoides (o PAM) si permite hacer esto.

B: Una matriz de disimilitud no permite calcular la matriz de varianzas y covarianzas con las que se construyen los segmentos.

Feedback: K-medias no utiliza la matriz de varianzas y covarianzas.

C: El algoritmo podría ser modificado para que se utilice la mediana de los segmentos y de este modo se pueda utilizar la matriz de disimilitud como input del modelo.

Feedback: Aunque esto es cierto, el algoritmo sería otro, en este caso K-Medoides.

D: Porque la matriz de distancias es una simplificación de los datos y por ende no contiene la información de todas las variables con las que se desea hacer la segmentación.

Feedback: La matriz de distancias sí contiene información de todas las variables.

**Question 3**

¿Se puede usar variables categóricas como inputs en la segmentación por K-Medias?

A: Sí, siempre y cuando las variables categóricas se conviertan en variables dicótomas para poder computar la distancia Euclideana.

Feedback: El algoritmo de K-Medias no puede ser aplicado a variables categóricas puesto que estas variables son discretas y no tienen un origen natural. Por ende, calcular la distancia Euclideana con estas variables no tiene un significado relevante.

B: Sí, siempre y cuando se cambie la fórmula para calcular las distancias entre observaciones por una que tenga en cuenta el uso de variables categóricas, como la distancia de Gower por ejemplo.

Feedback: El algoritmo de K-Medias solo puede implementarse haciendo uso de la distancia Euclideana por las definiciones subyacentes que tiene el algoritmo. K-Medias se basa implícitamente en distancias euclidianas por pares entre puntos de datos porque la suma de las desviaciones al cuadrado del centroide es igual a la suma de las distancias euclidianas al cuadrado por pares dividida por el número de puntos. De hecho, el término "centroide" es en sí mismo parte da la geometría euclidiana y hace alusión a la media multivariada. En general, las distancias no euclideanas no se ajustan bien al espacio euclideano constitutivo de este algoritmo.

\*C: No, no se puede computar la distancia Euclideana de las variables categóricas.

Feedback: Correcto, para hacer uso de variables categóricas se puede usar el algoritmo de K-Modas si solo se tiene como input este tipo de variables. En caso de contar con tipos de variables tanto continuas como categóricas se puede usar el algoritmo K-Prototipos el cual es una combinación convexa de K-Medias con K-Modas.

**Question 4**

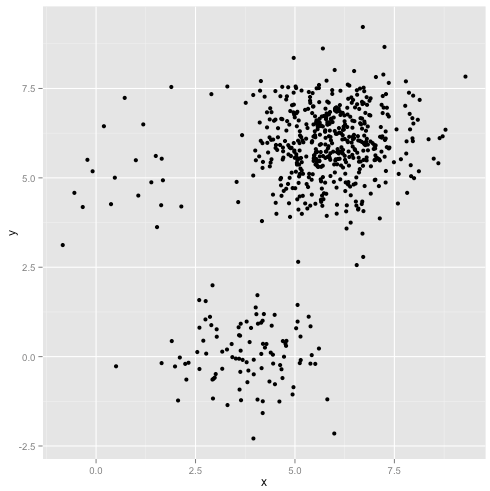
Seleccione los principales problemas que tiene el algoritmo de K-Medias (Más de una opción válida):

A: El algoritmo de K-Medias no garantiza convergencia.

Feedback: Aunque el algoritmo no es determinístico, este siempre logra converger a una solución.

\*B. El algoritmo tiene problemas para identificar clústeres con diferentes tamaños y varianzas.

Feedback: En la búsqueda por minimizar la suma de cuadrados dentro del grupo, el algoritmo de K-medias da más "peso" a los grupos más grandes. En la práctica, eso significa que el algoritmo ubica los centroides lejos de los grupos más pequeños y termina usándolos para dividir agrupaciones más grandes. Ejemplo a continuación:

\*C: El algoritmo puede encontrar segmentaciones erradas cuando hay presencia de datos atípicos.

Feedback: Los centroides pueden estar sesgados porque la media es una medida sensible a los datos atípicos.

\*D: Tener muchas variables hace que el algoritmo no encuentre segmentaciones coherentes.

Feedback: A medida que aumenta el número de dimensiones, una medida de similitud basada en la distancia converge a un valor constante entre cualquier ejemplo dado. Se recomiende reducir la dimensionalidad mediante el uso de PCA en los datos de características o mediante el uso de "agrupamiento espectral".