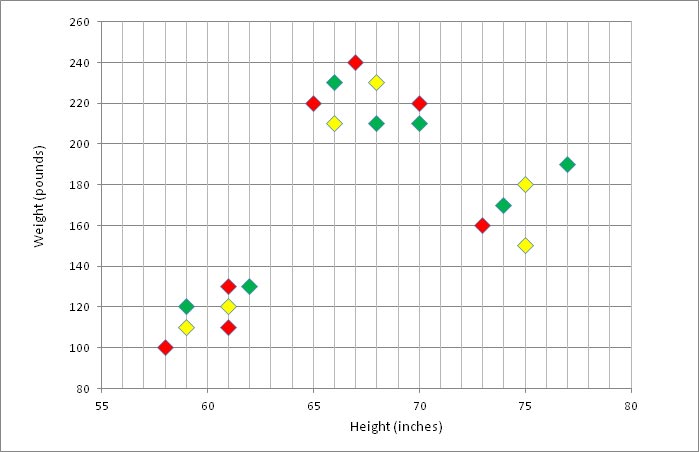
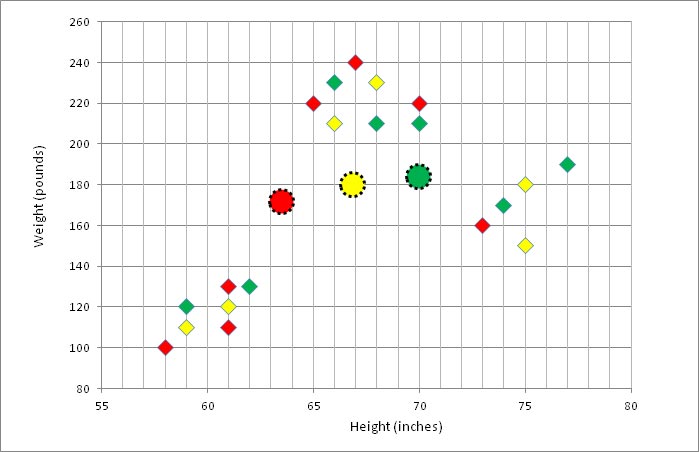
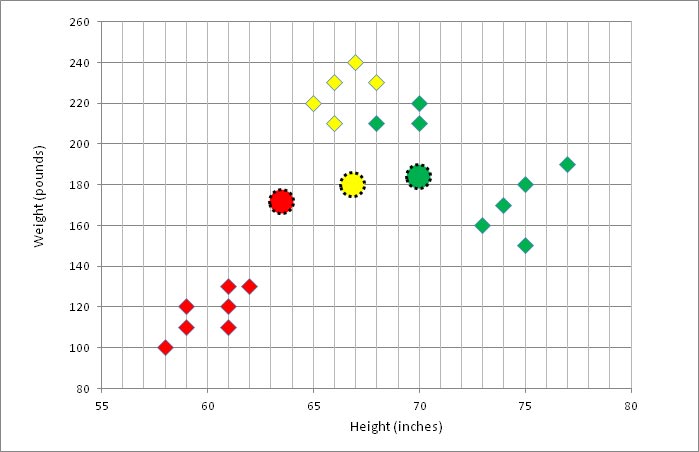
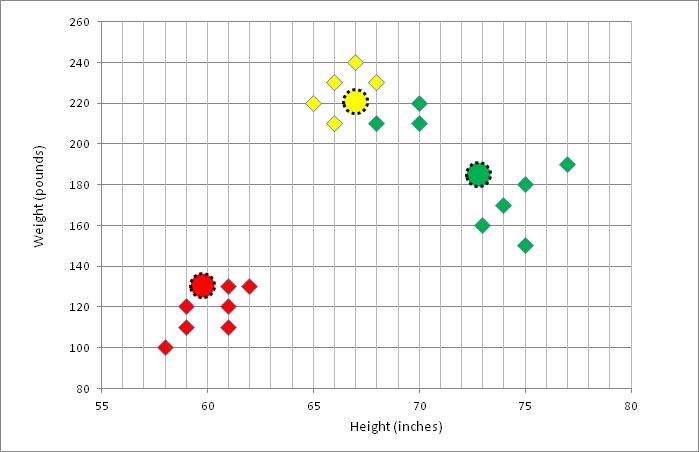
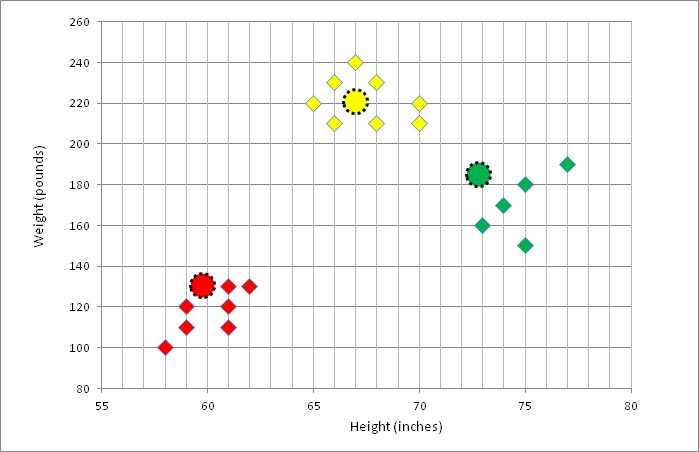
Comparacion algoritmos K-means y A priori

1. Algoritmo K-means  
   1. Definición  
      Método de análisis de conglomerados que apunta a dividir n observaciones en k conglomerados en los que cada observación pertenece a un conglomerado con la media más cercana. Cabe resaltar que el algoritmo solo es aplicable para datos puramente numéricos.
   2. Complejidad computacional  
      - NP-duro en el espacio euclidiano general d  
      - NP-duro para un número general de grupos k

- O(ndk+1) si d y k son fijos, n representa el número de entidades a agrupar

* 1. Problemas que resuelve  
     - La minimización del error de suma de cuadrados (SSE) en el análisis de regresión  
     - La segmentación de imágenes gráficas  
     - El desarrollo de motores de recomendación  
     - El mantenimiento de funciones de software eficientes que brindan seguridad pública en todo el mundo
  2. Normalizar los datos  
     Se deben normalizar los datos para que no existan valores atípicos que generen ruido en el análisis.
  3. Pasos:  
     a) Coloque k puntos en el espacio representado por los objetos que se agrupan. Estos puntos representan los centroides del grupo inicial.  
     b) Asignar cada objeto al grupo que tenga el centroide más cercano.  
     c) Cuando todos los objetos hayan sido asignados se debe recalcular la posición de los centroides.  
     d) Repetir el paso a y b hasta que la posición de los centroides no se modifique. Esto produce una separación de los objetos en grupos a partir de los cuales se puede calcular la métrica a minimizar.
  4. Ejemplo  
       
     - Cada punto de datos está coloreado en rojo, amarillo o verde para indicar la pertenencia al grupo.  
       
       
       
     - Se calcula el centroide de cada grupo, se indican mediante los círculos de colores.  
     

- Cada tupla de datos se examina y asigna al grupo más cercano representado por su media asociada.  
  
  
- Se computan los nuevos centroides del kluster  
  
  
- Agrupamiento estabilizado  


1. Algoritmo A priori  
   1. Definición  
      Algoritmo para la extracción frecuente de conjuntos de elementos y el aprendizaje de reglas de asociación a través de bases de datos transaccionales. Identifica los elementos individuales frecuentes en la base de datos y los extiende a conjuntos de elementos cada vez más grandes, siempre y cuando estos conjuntos de elementos aparezcan con suficiente frecuencia en la base de datos.
   2. Complejidad  
      La complejidad de este algoritmo se basa en la cantidad de elementos de los cuales se componen las transacciones como son los ítems.
   3. Aplicabilidad  
      A priori está diseñado para operar en bases de datos que contienen transacciones:

* Colecciones de artículos comprados por clientes
* Detalles de una frecuencia de visitas a sitios web
* Análisis de la cesta de supermercado
* Direcciones IP  
  1. ¿Cómo funciona?

Cada transacción es vista como un conjunto de elementos y dado un umbral *C* el presente algoritmo identifica los conjuntos de elementos que son subconjuntos de al menos *C* transacciones en la base de datos.

* 1. Pasos:  
     a) Emplea la “búsqueda en primer lugar” y una estructura de árbol hash para contar los conjuntos de elementos candidatos de manera eficiente.  
     b) Genera conjuntos de elementos candidatos de longitud *K* a partir de conjuntos de elementos de longitud *K-1.*c) Podar los candidatos que tienen un sub patrón poco frecuente.  
     d) El conjunto de candidatos contiene todos los frecuentes *K* conjuntos de elementos de longitud.  
     e) Escanear la base de datos de transacciones para determinar conjuntos de elementos frecuentes entre los candidatos.

|  |
| --- |
| **Conjuntos de artículos** |
| {1,2,3,4} |
| {1,2,4} |
| {1,2} |
| {2,3,4} |
| {2,3} |
| {3,4} |
| {2,4} |

* 1. Ejemplo

|  |  |
| --- | --- |
| **ít** | **Apoyo** |
| {1} | 3 |
| {2} | 6 |
| {3} | 4 |
| {4} | 5 |

|  |  |
| --- | --- |
| **ít** | **Apoyo** |
| {1,2} | 3 |
| {1,3} | 1 |
| {1,4} | 2 |
| {2,3} | 3 |
| {2,4} | 4 |
| {3,4} | 3 |

|  |  |
| --- | --- |
| **ít** | **Apoyo** |
| {2,3,4} | 2 |

Umbral: Un conjunto de elementos en frecuente si aparece en al menos 3 transacciones de la base de datos.  
  
- Número de ocurrencias

-No hay trillizos frecuentes. {2,3,4} está por debajo del umbral, y los demás trillizos se excluyeron porque eran súper conjuntos de pares que ya estaban por debajo del umbral.  
  
- Se han determinado los conjuntos frecuentes de elementos en la base de datos y se ha ilustrado como algunos elementos no se contaron porque ya se sabía que uno de sus subconjuntos estaba por debajo del umbral.

1. Bibliografía  
   <http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html>  
   <https://en.wikipedia.org/wiki/K-means_clustering>  
   <https://visualstudiomagazine.com/Articles/2013/12/01/K-Means-Data-Clustering-Using-C.aspx?Page=1>  
   <http://crsouza.com/2010/10/05/k-means-clustering-in-c/>  
   <https://en.wikipedia.org/wiki/Apriori_algorithm>