**La práctica 2 consistirá en un ejercicio (evaluable) previo a la práctica 3 ( dedicada más extensamente al clustering). La idea es realizarlo en una sesión de clase práctica o poco más**

Se recomienda ir guardando todos los resultados intermedios para luego poder componer el archivo final a enviar.

1-Trabajar con el archivo amValencia1.csv. Abrir y procesar con weka explorer. Existe también una versión en formato Excel, pero usaremos la de formato csv.La base de datos contiene información socio-económica ( de hace ya bastantes años) de los 59 municipios que constituyen el Área Metropolitana de Valencia ( según la delimitación de M.de Lejarza y M. Lejarza de 2001).La idea es familiarizarnos con distintos procedimientos de agrupación ( clustering) , mediante weka y R, con le objetivo de agrupar la municipios en grupos homogéneos

2- Eliminar los atributos Implantación y cercanía El archivo csv tiene una fila vacía de más ( que el archivo xlsx no tiene) Hay que eliminarla nada más abrir este archivo. Para ello editarlo clicando en el botón edit de explorer y eliminar la última fila.

3- Todavía en Explorer , hacer class🡪 no class (botón ventana de la derecha del explorer).Obtener ( con weka) una agrupación de **cuatro** grupos por el método farthest first. Ignorar el atributo muni [será necesario en todas las agrupaciones que hagamos con weka].Visualizar la agrupación obtenida ( tomar nota de aspectos notorios)

4- Añadir un atributo ( variable) que consigne la pertenencia de cada municipio a cada uno de estos cuatro grupos : weka.filters.unsupervised.attribute.AddCluster -W "weka.clusterers.FarthestFirst -N 4 -S 1" -I 1. Cuidado con las opciones del filtro: clusterer, distancia,número de grupos,atributos a excluir.Despues cambiar el nombre del atributo obtenido (cluster) a otro más ilustrativo clusterFF, por ejemplo. weka.filters.unsupervised.attribute.RenameAttribute -find ([\s\S]+) -replace clusterFF -R last ( Cuidado con las opciones del filtro)

5- Obtener una agrupación jerárquica ( bien encadenamiento simple o bien encadenamiento completo ) de cuatro grupos y también guardar la solución ( asignación). Será necesario, aquí, ignorar , además del atributo muni, los de los clusters obtenidos por otros métodos antes. Comprobar que para obtener 4 grupos por un método jerárquico hay que pedirle a weka que obtenga 5 (uno más). Tener en cuenta lo dicho en 4 para la obtención del atributo cluster de pertenencia.

6- Realizar lo mismo que en 5 pero para una agrupación por el método de Ward. Proceder de forma parecida en las opciones de ignorar atributos y solicitar número de grupos y al añadir el atributo cluster de pertenencia.

7- Guardar el nuevo archivo (sin implatanción, ni cercanía, y con tres nuevas variables de agrupación con formato csv

8- Abrir este nuevo archivo con R y llevar a cabo una agrupación jerárquica de 4 grupos que será la no realizada en el punto 5.

9- Comentar las coincidencias y diferencias entre las 4 agrupaciones obtenidas a partir del listado de los municipios y sus 4 adscripciones.

( antes de llevar a cabo el punto 10 se recomienda visualizar los valores medios de las distintas variables en cada uno de los grupos de cada una de las agrupaciones para identificar las variables “presuntamente relevantes” )

10- Interpretar el significado de las cuatro agrupaciones a partir de distintos análisis de la varianza ( comparaciones múltiples incluidas) de las variables más relevantes según las distintas agrupaciones ( consideradas como factores). OJO :Interpretar la agrupación NO ES determinar qué variables son más relavantes. El punto de vista del análisis no está puesto en las variables sino en los grupos.

11- Componer un informe final que incluya todo el trabajo realizado en un archivo con formato de Word o pdf y enviar .

Diccionario de variables:

Nortesur : al norte ( 1) al sur (-1) de la ciudad de Valencia ( Valencia toma el valor 0 )

idependencia =

=100\*( personas menores de 15 + personas mayores de 65) / personas entre 16 y 64

Tnatalidad : nacimientos por 1000 hab

Tmortalidad: defunciones por 1000 hab

Tcrecimiento: natalidad - mortalidad

Tmigracion: Inmigrantes-emigrantes por 1000 hab

Envejecimiento: mayores de 65 / menores de 65, en %

Nlineastelef : líneas telefónicas FIJAS por 1000 hab

Nveh : vehículos CENSADOS en el municipio por cada 1000 hab

Indimas : ( varones / mujeres) \* 100

Segvivienda : porcentaje de viviendas censadas destinadas a segunda residencia

Vivdesocu: porcentaje de viviendas censadas que están desocupadas

Ta : tasa de actividad

Taf : tasa de actividad femenina

Iacomercia/ iafin / iindus : Número de locales de actividad de estos sectores por ha de superficie urbana

Cenerdomest: consumo de energía (eléctrica) doméstica

Cenergindus: consumo de energía eléctrica industrial

Formacad : índice de formación académica de la población

Iparlabser/iparlabindustria/ iparlabconstr ( Indice de partición laboral en cada sector)