

KNN:

Uso: Agrupar datos según criterio del propio algoritmo, sin necesidad de indicarle ni siquiera el número de cluster a utilizar.

Descripción: Se basa en el reconocimiento de patrones de vecindad.

Para ellos, va calculando las distancias que separan a cada uno de los elementos y les va agrupando por el que está más cercano.

Para el cálculo de distancias se suele utilizar la euclídea



#### Cluster Jerarquico:

Generan grupos en cada una de las fases del proceso buscando un numero de cluster que hacen la agrupacion activa, es capaz de fijar por si solo el numero de cluster por ello se puede utilizar de forma exploratoria para ver que numero de grupos se pueden existir

Cuando tengamos eso podemos utilizar el no jerarquico cuando el numero de cluster esta fijado

Los metodos jerarquicos se pueden dividir segun su estrategia tanto en aglomerativos como divisivos.

**Agglomerativos** : Empieza con todos los elementos disgregados y va juntandolos poco a poco para arriba.

**Divisivos**: Hace lo contrario tiene los elementos juntos y va haciendo particiones como grupos queramos, va bajando en elarbol

Que metodos tenemos en el analisis cluster?

Metodo vecino próximo  
Metodo vecino lejano  
Agrupación de Vinculación promedio  
Varianza minima  
Centróide  
Mediana

#### Agglomerativos:

- Vecino mas próximo o enlace simple
- Vecino mas lejano o enlace completo
- Método de agrupación de vinculación promedio

#### Divisivos:

- Metodo de agrupación de vinculación promedio
- Metodo de Ward o Varianza minima
- Metodo de la Mediana

#### Cluster No Jerarquico:

Generan grupos en cada una de las fases del proceso buscando un numero de cluster que hacen la agrupacion activa, es capaz de fijar por si solo el numero de cluster por ello se puede utilizar de forma exploratoria para ver que numero de grupos se pueden existir

Cuando tengamos eso podemos utilizar el no jerarquico cuando el numero de cluster esta fijado

Los metodos jerarquicos se pueden dividir segun su estrategia tanto en aglomerativos como divisivos.

aglomerativos: Empieza con todos los elementos disgregados y va juntandolos poco a poco para arriba.

divisivos: Hace lo contrario tiene los elementos juntos y va haciendo particiones como grupos queramos, va bajando en elarbol

Que metodos tenemos en el analisis cluster?

Metodo vecino próximo  
Metodo vecino lejano  
Agrupación de Vinculación promedio  
Varianza minima  
Centróide  
Mediana

#### K-Means

Uso: Agrupar datos segun criterio del propio algoritmo, necesita indicarle el numero de cluster que se desea obtener.

Minimiza la distancia entre los elementos del cluster a la vez que maximiza la distancia entre los grupos inter cluster

Descripción: Metodo de clasificaciomm que busca minimizar la distancia entre grupos y maximizar la distancia intra-grupos.

Para ello, el algoritmo identificara k centroides los cuales comparan su valor medio con el de los datos restantes seleccionando aquellos mas cercanos para realizar los clusters o agrupamientos

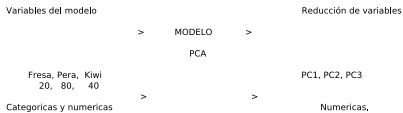


#### Análisis de componentes principales

Uso reducir el numero de variables del conjunto original utilizando sus componentes principales

Descripción: Método de reducciomm de componentes que se utiliza para analizar las interrelaciones que existen entre un numero elevado de variables metricas explicando dichas interrelaciones en terminos de un numero menor de variables denominadas componentes principales (siempre y cuando no seann observables)

Los nuevos componentes principales seran combinacion lineal de las variables originales y seran independientes entre si,

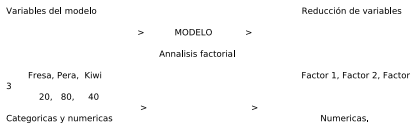


#### Análisis factorial

Metodo para reducir el numero de variables del conjunto original utilizando sus factores

Descripción: El método factorial o de reducción de variables que se utiliza para analizar las interrelaciones que existen entre un numero elevado de variables metricas explicando dichas interrelaciones en terminos de un numero menor de variables denominadas factores (siempre y cuando sean inobservables)

Los factores seran combinacion lineal de las variables originales y seran independientes entre si,

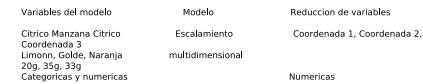


#### Escalamiento Multidimensional

Uso: Reducir un numero de variables del conjunto original utilizando una escala multidimensional

Descripción: Método de reduccion de variables que se utiliza para analizar las interrelaciones que existen entre un numero elevado de variables metricas explicando dichas interrelaciones en terminos de un numero menor de variables denominadas coordenadas

Las nuevas coordenadas seran combinacion lineal de las variables originales y seran independientes entre si,



#### Análisis cluster

Clasificar en grupos atendiendo a las variables de los datos.  
Ejemplo: Buscar grupos de clientes con habitos de compra similares,

#### Aprendizaje NO supervisado.

Técnicas que permiten inferir modelos para extraer conocimiento de conjuntos de datos donde a priori se desconoce

Estas se pueden aplicar sin necesidad de tener los datos etiquetados para el entrenamiento.

**Al disponer de los datos de entrada solo podemos predecir una estructura de los datos** y con ello intentar encontrar algun tipo de organización que simplifique el análisis, por lo que tiene carácter exploratorio.

No necesitamos saber como son esos datos

#### Reducción de dimensionalidad:

Se trata de reducir el numero de variables a tener en cuenta en el análisis

Ejemplo: Si tenemos un modelos de predicción meteorológica podemos eliminar una variable que nos indique si ha llovido no en caso de que dispongamos de otra variable que nos de el volumen de lluvia.