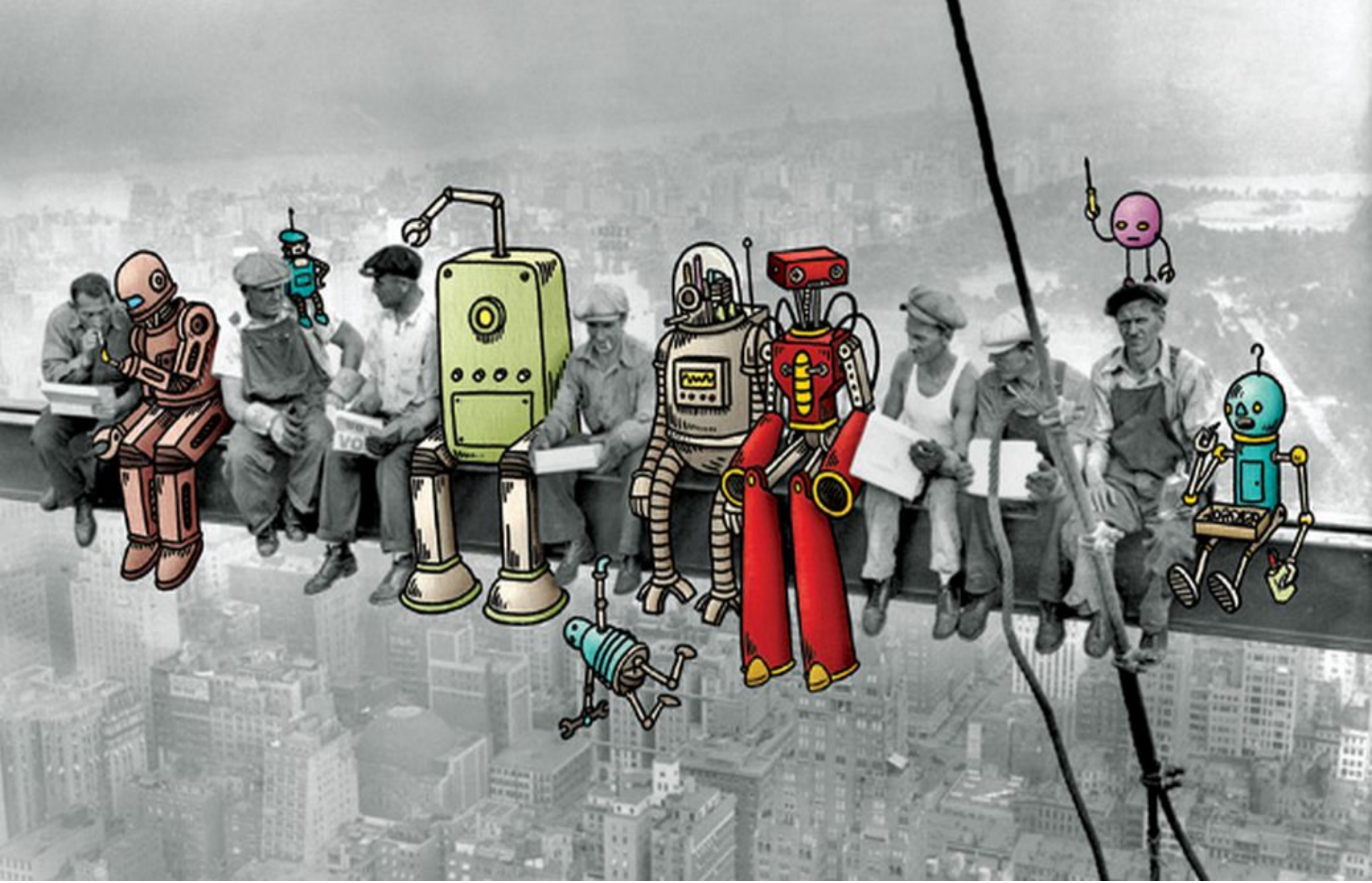


1. Métodos de Aprendizaje No Supervisado - Descripción

DP01 - Machine Learning (II)

DP - Data Proficiency - Data Analytics Journey

DP01 - Machine Learning (II)

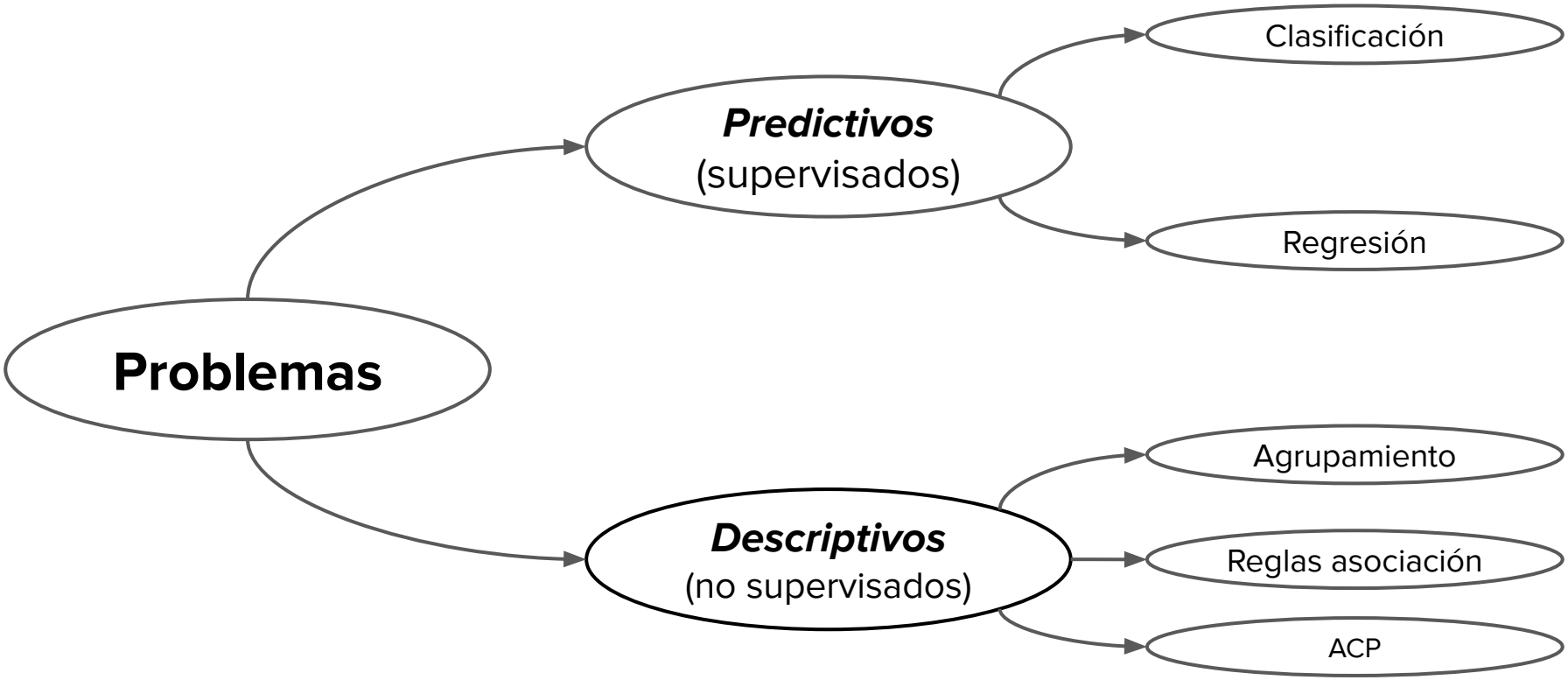
1. Métodos de Aprendizaje No Supervisado - Descripción

- **Objetivos de aprendizaje**
  - Entender la diferencia que hay entre las técnicas de aprendizaje no supervisado y las supervisadas
  - Conocer y aplicar la técnica de agrupación de clustering
  - Conocer y aplicar la técnica de reglas de asociación
  - Conocer y aplicar las diferentes técnicas de reducción de la dimensionalidad

1. Métodos de Aprendizaje No Supervisado - Descripción

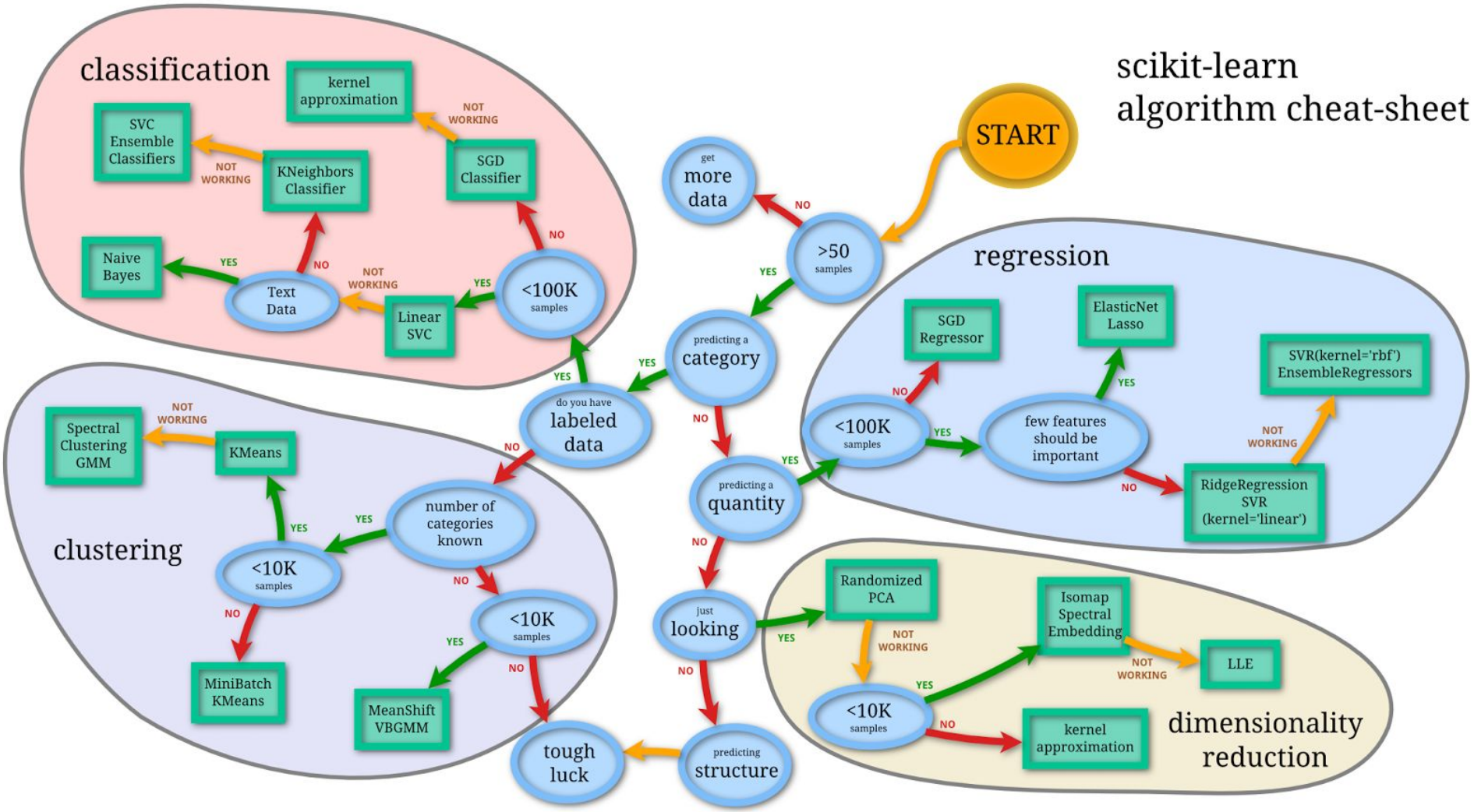
1. Métodos de Aprendizaje No Supervisado - Descripción

Tipos de problemas a resolver



1. Métodos de Aprendizaje No Supervisado - Descripción

Tipos de problemas a resolver (II)



1. Métodos de Aprendizaje No Supervisado - Descripción

Tipos de problemas a resolver (III)

- **Supervisados (Predictivos)**
  - Cuando disponemos del valor que quisiéramos que nuestro modelo diera ante una determinada entrada
  - Tenemos datos “etiquetados”
    - Con la clase deseada o Con el valor esperado
  - Realizan predicciones del valor de salida a partir de datos

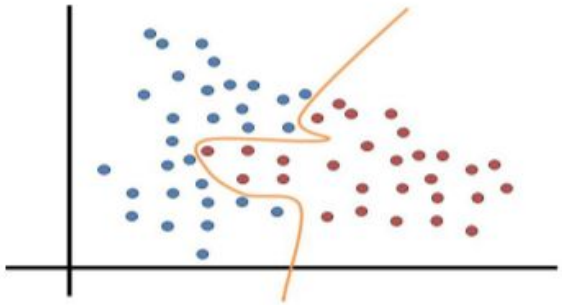
Deuda	Salario	Moroso
100.000	10.000	SI
110.000	30.000	NO
80.000	50.000	NO
90.000	45.000	NO

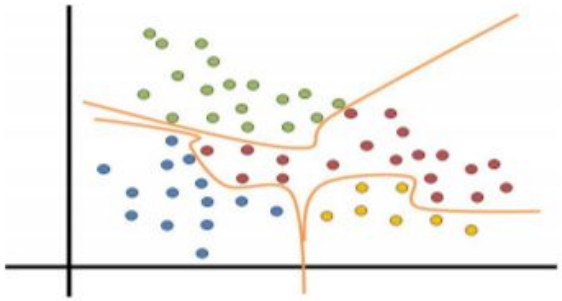
Salario	Edad	Préstamo
10.000	25	100.000
30.000	50	110.000
50.000	45	20.000
45.000	27	250.000

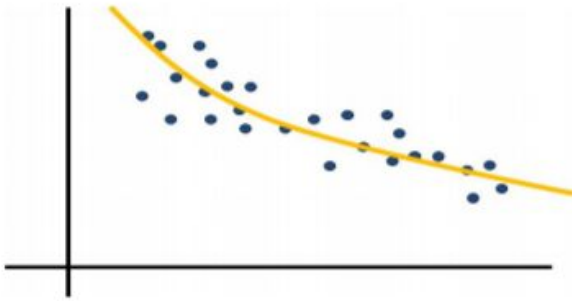
1. Métodos de Aprendizaje No Supervisado - Descripción

Tipos de problemas a resolver (IV)

- **Supervisados (Predictivos)**
  - Clasificación: Cuando la variable a predecir es una categoría.
    - Binaria: {Sí, No}, {Azul, Rojo}, {Fuga, No Fuga}...
    - Múltiple: {Comprará Producto1, Producto2...}...
    - Ordenada: {Riesgo Bajo, Medio, Alto}...
  - Regresión: Cuando la variable a predecir es una cantidad
    - Precio, cantidad, tiempo, etc.







1. Métodos de Aprendizaje No Supervisado - Descripción

Tipos de problemas a resolver (V)

Para más detalle de esta introducción, podéis consultar el curso DA05 del nivel Advanced, donde hay una introducción al Machine learning

- **No Supervisados (Descriptivos)**
  - Los usaremos cuando no disponemos del valor que quisiéramos que nuestro modelo diera ante una determinada entrada
  - Su objetivo es modelar y describir la estructura o distribución interna de los datos
  - Muchas aplicaciones reales hacen uso de estos datos
    - Hay más datos, y son baratos
    - Etiquetarlos puede ser costoso
    - Fáciles de obtener
  - Por ejemplo:
    - **Agrupamiento - Clustering**: Buscan encontrar grupos dentro de los datos de elementos similares ➔ permite así crear perfiles
      - Clientes con hábitos de compra similares
      - Productos vendidos en fechas similares
    - **Asociación**: buscan reglas que describen la mayor parte posible de los datos de los que se disponen ➔ permite crear reglas de co-ocurrencia de valores de variables
      - Productos que se compran juntos

1. Métodos de Aprendizaje No Supervisado - Descripción
Introducción

- Son métodos de aprendizaje que buscan reconocer patrones para poder etiquetar las nuevas entradas
- Dadas las observaciones de un concepto (que incluso en muchas ocasiones es desconocido), el objetivo es obtener la caracterización de ese concepto
- Hay dos **aproximaciones principales**:
  - **(1) Agrupamiento**
    - El objetivo es encontrar grupos que reflejen la estructura del espacio de entrada
    - Destaca fundamentalmente el Clustering
  - **(2) Descubrimiento**
    - El objetivo es encontrar una serie de leyes o funciones que describan las observaciones
    - Destacan el análisis de Reglas de Asociación y la familia de técnicas para reducir la dimensionalidad (Análisis Factorial)

1. Métodos de Aprendizaje No Supervisado - Descripción

Análisis multivariante

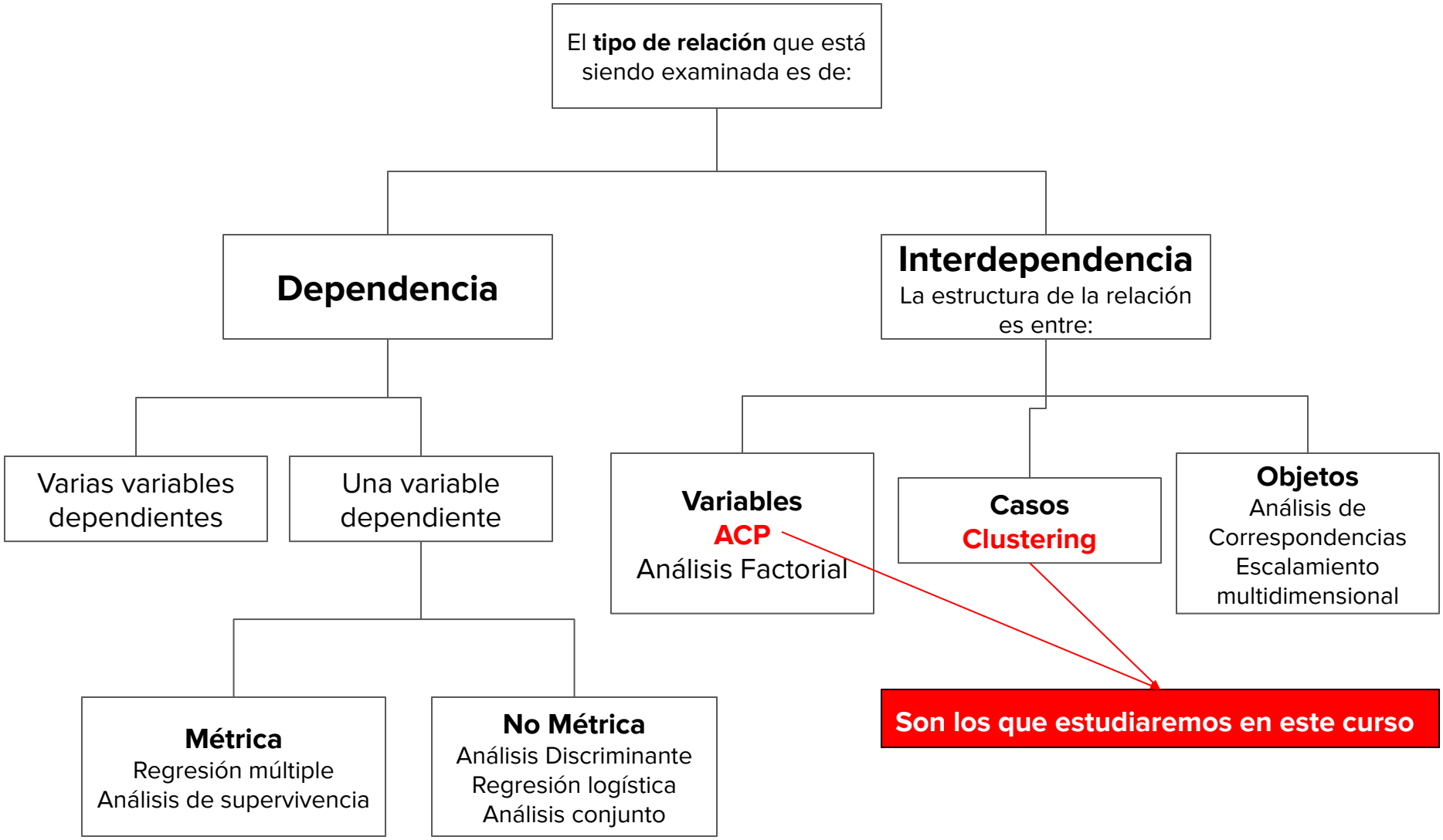
- Las técnicas de aprendizaje no supervisado forman parte también y se entienden con los métodos de análisis multivariante
- Éstos tienen como finalidad analizar simultáneamente conjuntos de datos multivariantes
  - En el sentido de que hay varias variables medidas para cada individuo u objeto estudiado
- Su razón de ser radica en un mejor entendimiento del fenómeno objeto de estudio obteniendo así información que los métodos univariantes y bivariantes no pueden proporcionar

1. Métodos de Aprendizaje No Supervisado - Descripción
Análisis multivariante (II)

- Son técnicas que se diferencian unas de otras según su área de aplicación; según se refiera a uno o más problemas y según se requiera uno o más grupo de variables
- Se pueden clasificar en tres grandes grupos:
  - **(1) Métodos de dependencia**
    - Suponen que las variables analizadas están divididas en dos grupos: las variables dependientes y las variables independientes. El objetivo de los métodos de dependencia consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma.
  - **(2) Métodos de interdependencia**
    - Estos métodos no distinguen entre variables dependientes e independientes y su objetivo consiste en identificar qué variables están relacionadas, cómo lo están y por qué.
  - **(3) Métodos estructurales**
    - Suponen que las variables están divididas en dos grupos: el de las variables dependientes y el de las independientes. El objetivo de estos métodos es analizar, no sólo como las variables independientes afectan a las variables dependientes, sino también cómo están relacionadas las variables de los dos grupos entre sí

1. Métodos de Aprendizaje No Supervisado - Descripción

Análisis multivariante (III)



1. Métodos de Aprendizaje No Supervisado - Descripción

Fases de un proyecto

1. Métodos de Aprendizaje No Supervisado - Descripción
Aplicaciones

- **Medicina**
  - Evaluar la presencia o ausencia de determinados síntomas clínicos para diagnosticar la enfermedad de un paciente (análisis discriminante)
  - Para estimar la probabilidad de que la sintomatología de una determinada enfermedad reaparezca antes de un período determinado, conocidos el tiempo de respuesta al tratamiento y los distintos hábitos del paciente, (Regresión logística).
  - Se tabula las frecuencias de ciertos estímulos y sus respuestas. Interesa obtener una representación bidimensional de las correspondencias entre estímulos y respuestas (Análisis Factorial de Correspondencia)

1. Métodos de Aprendizaje No Supervisado - Descripción
Aplicaciones (II)

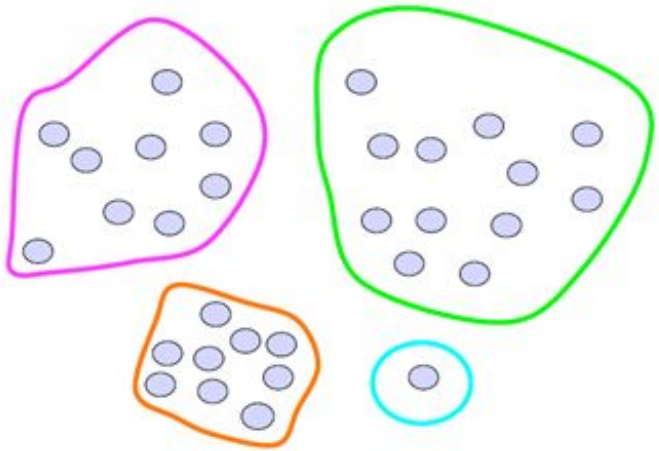
- **Biología**
  - Se miden diferentes variables biométricas en los individuos de una misma especie. Se desea detectar componentes de tamaño y forma (Análisis de Componentes Principales).
  - Las observaciones de “p” variables biométricas representativas de los individuos de una especie, se obtienen para estudiar la variabilidad entre diferentes especies o razas geográficas (Análisis Canónicos).

1. Métodos de Aprendizaje No Supervisado - Descripción
Aplicaciones (III)

- **Sociología**
  - Con referencia a determinadas características sociales, políticas y geográficas se mide la similaridad de un grupo de naciones. (Escalamiento Multidimensional)
- **Psicología**
  - Los resultados de un test de inteligencia de “n” ítems basados en una muestra. Para detectar los factores de la inteligencia (Análisis Factorial).
- **Investigación de Mercados**
  - Se quiere determinar los beneficios subyacentes que buscan los consumidores en la compra de una pasta dental. (Análisis Factorial).
  - Para el análisis de percepciones y preferencias del consumidor (Escalamiento Multidimensional).

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering

- Los métodos de clustering buscan agrupar un conjunto de datos en “clusters” (grupos), según cierta medida de distancia
  - Datos dentro del mismo cluster deben estar cerca los unos de los otros
  - Datos de clústeres diferentes deben estar lejos los unos de los otros
- Un cluster es una agrupación de muestras de datos
  - Similares a aquellas pertenecientes al grupo
  - Diferentes de aquellas no pertenecientes al grupo
- Análisis cluster
  - Comprender la estructura de los datos
  - Encontrar similitudes entre datos
  - Agrupar elementos
  - Es no supervisado ➜ no hay “clases” definidas



1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering (II)

- Conjunto de técnicas multivariantes que agrupan objetos (o variables) en grupos o clusters
- Análisis descriptivo, ateórico y no inferencial.
- Se utiliza para el descubrimiento de patrones de grupos de datos y no para la predicción
- No tiene bases estadísticas sobre las cuales deducir inferencias.
- Fuertes propiedades matemáticas (se basan en algoritmos), pero no fundamentos estadísticos.
- No hay supuestos previos.

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (III)

- **¿Cuántos grupos?**
  - Grupos o clusters no definidos a priori. Diferencia con los métodos supervisados.
- **¿Cómo buscarlos?**
  - Los objetos dentro de un cluster sean similares o cercanos entre sí en algún sentido (gran similaridad intra-clase) y diferentes o alejados a los objetos de otro cluster (baja similaridad inter-clase)
- Un ejemplo
  - Agrupar una baraja de cartas
- ¿Por tipo de carta?
  - ¿Por palos de la baraja?
  - ¿Por números de carta?
  - .....
- Agrupar valores en una ruleta:
  - ¿Por color?
  - ¿Pares e impares?
  - ¿Por números que acaben en el mismo dígito?
  - .....

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (IV)

- Algunas aplicaciones:
  - **Marketing**: segmentar clientes en grupos con características sociodemográficas o patrones de compra parecidos para hacer campañas de marketing dirigidas
  - **Política**: agrupar individuos con orientaciones políticas similares
  - **Text Mining**: agrupar documentos con temáticas parecidas
  - **Análisis de redes sociales**
  - **Finanzas**: agrupar valores de acciones para gestión de carteras de inversión
  - ...
- Y también para....
  - **Comprender los datos** ➔ el utilizar métodos de clustering nos permitirá ver si los datos están agrupados o no, con lo que ganaremos conocimiento sobre nuestros datos, que nos permitirá tomar decisiones más acertadas
  - **Etapa de preprocesamiento** ➔ para beneficiar en otra tarea del ciclo de vida de los datos.
    - ***Calidad de datos***: si hay un cluster de datos que son similares entre sí porque comparten cierto valor de atributo

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (V)

- **Preguntas a resolver en el análisis cluster**
  - ¿Qué son grupos naturales entre los objetos?
    - Definición de agrupamiento
    - Qué hace que los objetos estén relacionados
    - Definición de distancia-similaridad entre objetos
  - Representación de objetos
    - ¿qué variables?
    - ¿normalizadas?
  - ¿Cuántos clusters?
    - Fijados a priori
    - Los que digan los datos
    - Evitar clusters triviales demasiado grandes o pequeños

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (VI)

- Una buena agrupación está compuesta por clusters con
  - Alta similitud intra-clase
  - Baja similitud inter-clase
- Métricas de evaluación de la calidad
  - Medidas de similitud inter/intra cluster
  - Inspección manual
  - Comparación con “etiquetas” pre-diseñadas
- Así, el análisis cluster, se basa en minimizar/maximizar ciertas medidas de distancia entre las instancias del conjunto de datos
  - d(x,y)
    - Debe de ser mayor que cero
    - Es igual a cero si sólo si ambos elementos son iguales
    - Es simétrica (d(x,y)=d(y,x))
    - La distancia entre x e y es menor o igual que la distancia entre x y z más la distancia entre z e y (desigualdad triangular)

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (VII)

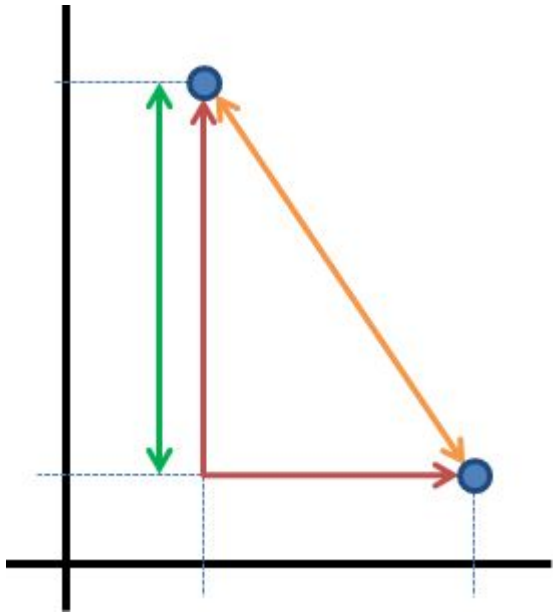
- Medidas de similaridad y distancia
  - **(1) Medidas de correlación**
    - Sirven para agrupar sobre todo variables continuas.
    - Elevadas correlaciones implican similitud y bajas correlaciones implican falta de ella.
  - **(2) Medidas de distancia**
    - Sirven para objetos u observaciones
      - Representan la similitud como la proximidad de las observaciones respecto a las otras.
    - Valores más elevados implican una menor similitud.
- Tipos principales de medidas de distancia:

$$d(a,b) = \left(\sum_{i=1}^{d} |a_i - b_i|^p\right)^{1/p}$$

  - **p=2 Distancia Euclídea**: Distancia en línea recta.
    - Problema de la escala de las variables (si cambias la escala de las variables cambian los grupos formados).
    - considerar la posibilidad de tipificar los datos.
  - **p=1 Distancia de Manhattan**: Distancia entre bloques o manzanas de ciudad.
  - **p= Infinito.** Máxima distancia

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (VIII)

- Las distancias más comunes son las euclídeas
  - **Norma L2**: raíz cuadrada de la suma de los cuadrados de las diferencias entre x e y en cada dimensión
    - La noción más común de distancia
  - **Norma L1**: suma de las diferencias en cada dimensión
    - Distancia de Manhattan: distancia si sólo te Mueves a través de las coordenadas
  - **Norma L∞**: máxima distancia de las diferencias entre x e y en cada dimensión



1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (IX)

- A quién se parece más Pedro?
  - (1) Pedro: Salario = 25.000, Edad= 30
  - (2) Juan: Salario = 27.000, Edad= 50
  - (3) Daniel: Salario = 20.000, Edad= 32
    - A Juan, que gana parecido pero es más senior.
    - A Daniel, que tiene una edad parecida pero cobra una cantidad menor
  - ¿Por qué?

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (X)

- ***Ejercicio***: si tenemos 2 clientes y 8 productos representados por estos datos, calcular la distancia euclídea entre ellos

	Producto 1	Producto 2	Producto3	Producto 4	Producto 5	Producto 6	Producto 7	Producto 8
Cliente 1	0	0	1	0	0	1	1	1
Cliente 2	0	1	0	1	0	0	1	1

  - ¿Qué está contando la distancia euclídea?
  - Si en vez de productos, las variables fueran idiomas hablados por los clientes, ¿tendría el mismo sentido contar las coincidencias 0-0 que las 1-1?

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (XI)

- Calcular la similitud entre los dos clientes del ejemplo anterior:
  - Considerando únicamente como valores coincidentes los 1-1
  - Considerando como valores coincidentes los 1-1 y 0-0

		Cliente2		
		1	0	total
Cliente1	1	a	b	a+b
	0	c	d	c+d
		a+c	b+d	p=a+b+c+d

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (XII)

- Para medir la distancia entre las instancias de datos, es necesario que todos los atributos estén en la misma escala
  - **Normalización**: escala los valores numéricos en el Rango [0,1]

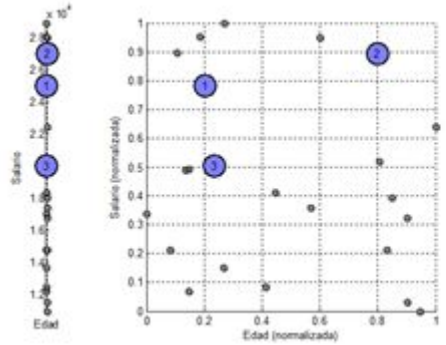
$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

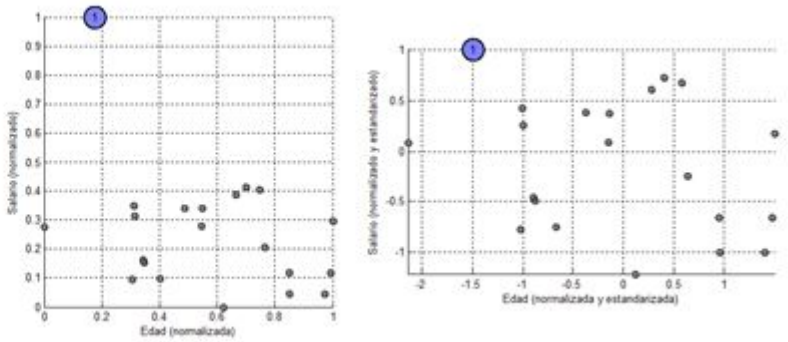
  - **Estandarización**: hace que la distribución de los datos sea normal
    - En el sentido estadístico de la palabra

$$x_{new} = \frac{x - \mu}{\sigma}$$

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (XIII)

- Tanto normalizar como estandarizar tienen sus ventajas e inconvenientes:
  - **Normalizar** nos asegura que los valores estarán entre 0 y 1, para todos los atributos
    - (Estandarizar, no)
  - **Estandarizar** mitiga el efecto negativo de los outliers, hace que los datos estén “mejor distribuidos”, lo que facilita la tarea de minería
    - (Normalizar, no)





1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (XIV)

- Aplicar ambos procesos, estandarizar y después normalizar
- Estudiar los outliers de los datos
  - Si no hay, podemos evitar estandarizar
- Hacer un proceso de limpieza de outliers previo
- Normalizar según rangos intercuartílicos
  - Ajustar valores que se salen de ellos

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering (XV)

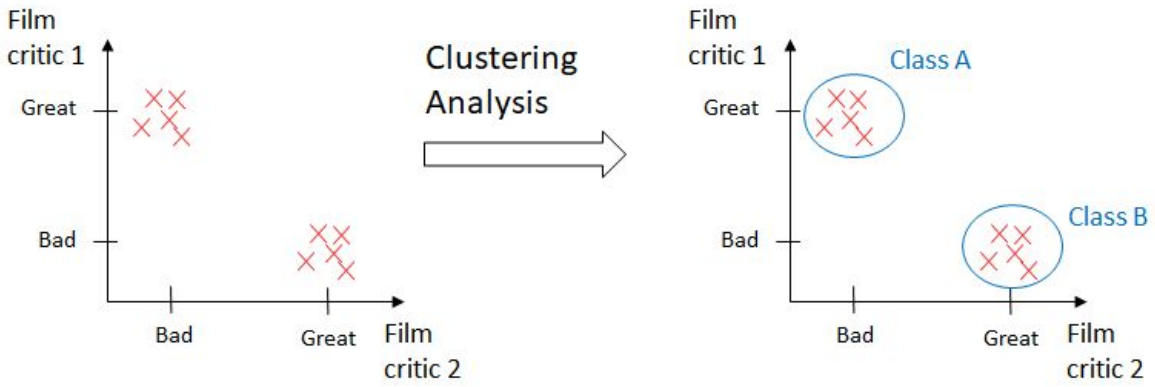
- Hamming: número de elementos diferentes
  - Útil para datasets con atributos categóricos
- Diferencia de ranking
  - Atributos ordinales
- Distancia de editado
  - Cuando los datos son strings
  - El menor número de inserciones y borrados (y modificaciones, en ciertas ocasiones) necesarios para pasar de la cadena x a la cadena y
  - (Inversa de) Longitud de la mayor subsecuencia común
- Matriz de distancias específica

	Ingeniero	Empresariales	Abogado
Ingeniero	0	0.5	1
Empresariales	0.25	0	0.25
Abogado	1	0.5	0

1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Tipos

- **(1) En función del número de clústeres**
  - **Clusters disjuntos**
    - Cada objeto pertenece a un único cluster
  - **Clusters coincidentes**
    - Cada objeto puede pertenecer simultáneamente a más de un cluster
- **(2) En función de cómo se construyen**
  - **Algoritmos particionales**
    - Empezar con una partición al azar y refinarla de manera recursiva
  - **Algoritmos jerárquicos**
    - Aglomerativos (bottom-up), top-down



1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - K-means

- Probablemente el más utilizado y conocido
- Asigna cada observación a uno de los k clusters
- K es un número definido a priori
- Minimizar las distancias intra cluster y maximizar las inter clase

K-means clustering is not a free lunch

I recently came across this question on Cross Validated, and I thought it offered a great opportunity to use R and ggplot2 to explore, in depth, the assumptions underlying the k-means algorithm. The question, and my response, follow.

K-means is a widely used method in cluster analysis. In my understanding, this method does NOT require ANY assumptions, i.e., give me a data set and a pre-specified number of clusters, k, then I just apply this algorithm which minimize the SSE, the within cluster square error.

So k-means, it is essentially an optimization problem.

I read some material about the drawback of k-means, most of them says that:

- *k-means assume the variance of the distribution of each attribute (variable) is spherical;*
- *all variables have the same variance;*
- *the prior probability for all k clusters are the same, i.e. each cluster has roughly equal number of observations; If any one of these 3 assumptions is violated, then k-means will fail.*

I could not understand the logic behind this statement. I think k-means method essentially makes no assumptions, it just minimizes the SSE, I cannot see the link between minimizing the SSE and those 3 "assumptions".

1. Métodos de Aprendizaje No Supervisado - Descripción

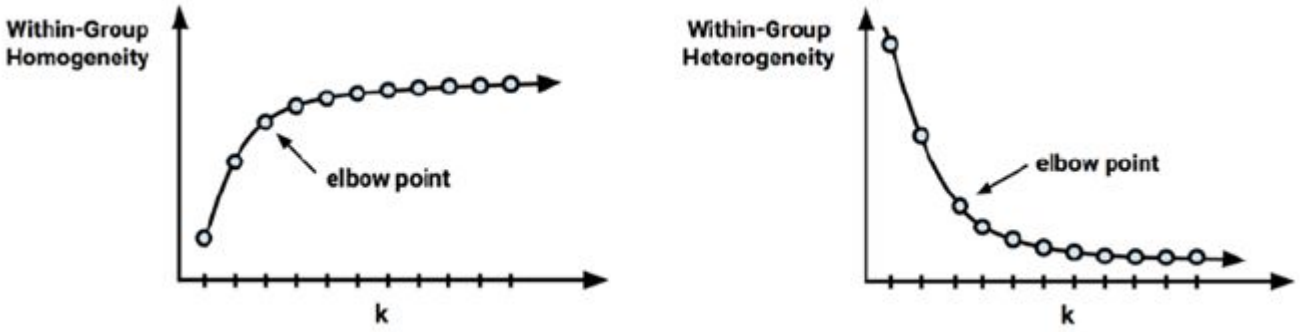
Clustering - K-means (II)

- ¿Cómo funciona el algoritmo?
  1. Elegir el valor de K (número de clusters).
  2. Elegir los centros de los k clusters, por ejemplo al azar (centroides)
  3. Asignar cada objeto al grupo más cercano (por ejemplo distancia euclídea)
  4. Re-estimar los centros de los k clusters, asumiendo que las asignaciones a los grupos están bien
  5. Repetir el paso 3 hasta que no haya más cambios
- Se puede cambiar el punto 2, empezando con k centroides iniciales
- La mayor parte de las reasignaciones ocurren en la primera iteración del algoritmo

1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering - K-means (III)

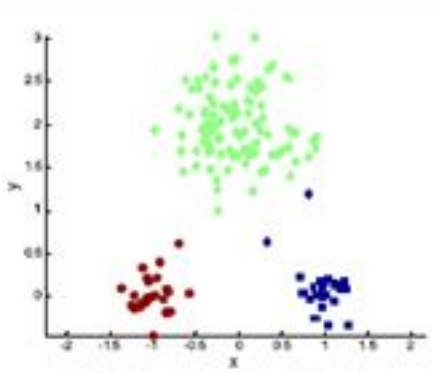
- Elegir el **número de clústers**
  - **Conocimiento a priori**: por ejemplo, si clasificamos películas, k=nº de géneros
  - **Dirigidos por el negocio**: por ejemplo, el departamento de Marketing sólo tiene recursos para hacer 3 campañas distintas de marketing
  - Sin nada de lo anterior: k=raíz(n/2)

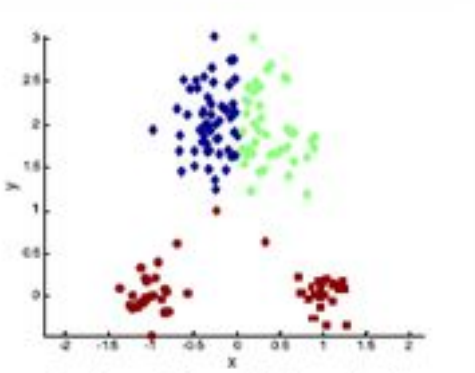
Regla del codo



1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering - K-means (IV)

- Se trata de un **método estocástico**
  - Según su implementación, los puntos iniciales se escogen con cierto factor de aleatoriedad, por lo que el resultado obtenido NO siempre es el mismo
- El método seleccionado para elegir los centroides iniciales es crítico para su desempeño
- Usar otro método para determinarlos
- Elegir un número mayor que k, y seleccionar entre ellos


Optimal Clustering


Sub-optimal Clustering

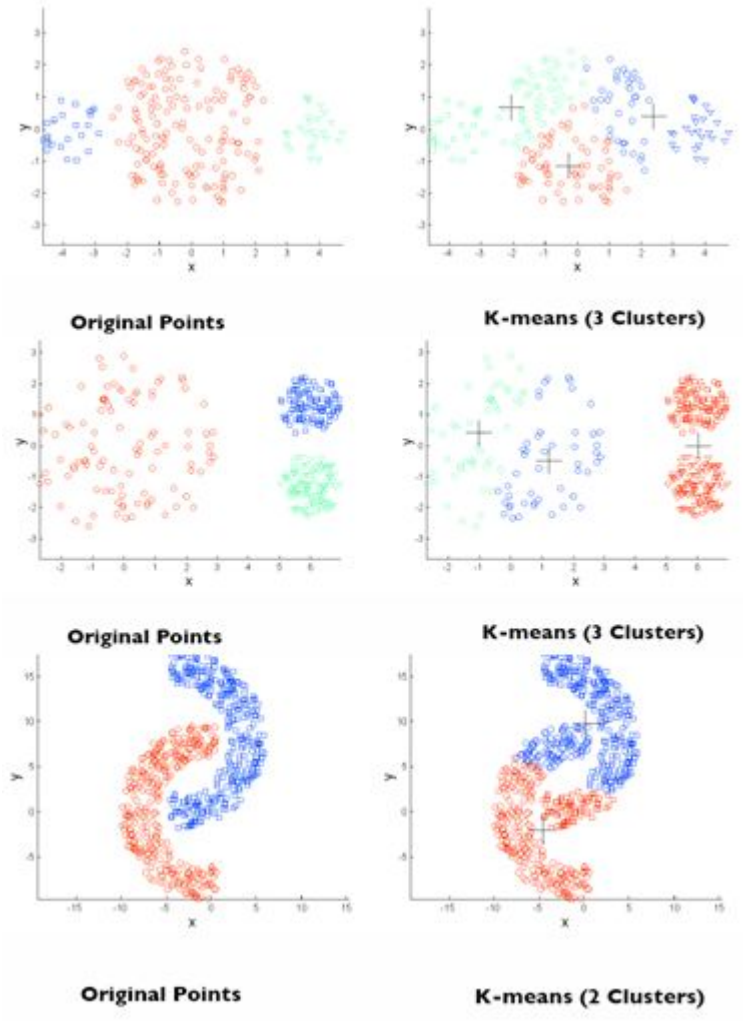
1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering - K-means (V)

- **Ventajas**
  - Principios no estadísticos
  - Muy flexible
  - Funciona bien en casos de la vida real
  - Rápido: no hay calcular las distancias entre todas y cada una de las observaciones
- **Desventajas**
  - No muy sofisticado
  - No está garantizado encontrar en número de clusters óptimo
  - Sensible a outliers que pueden formar clusters propios
  - La solución final depende del punto de partida

1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - K-means (VI)

- **Limitaciones**
  - Principalmente, su desempeño se ve mermado cuando los clusters tienen
    - Diferentes tamaños
    - Diferentes densidades
    - Formas no globulares
  - (Al igual que casi todos) También presenta problemas cuando los datos contienen outliers
  - Una solución puede ser hacer un número superior de clusters, y luego “unir las partes”



1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering - K-means (VII)

- Es un método simple
- Se debe seleccionar el número de clusters a priori
- Sensible a outliers
- Sus variantes giran en torno a:
  - Elección de los elementos iniciales
  - Cálculo de distancias
  - Uso de diferentes definiciones de centroide (más allá de la media)
- Extensiones
  - Muchas... BFR es la más conocida:
    - Especialmente diseñada para lidiar con grandes volúmenes de datos
    - Mantiene un resúmen estadístico de los datos ya procesados

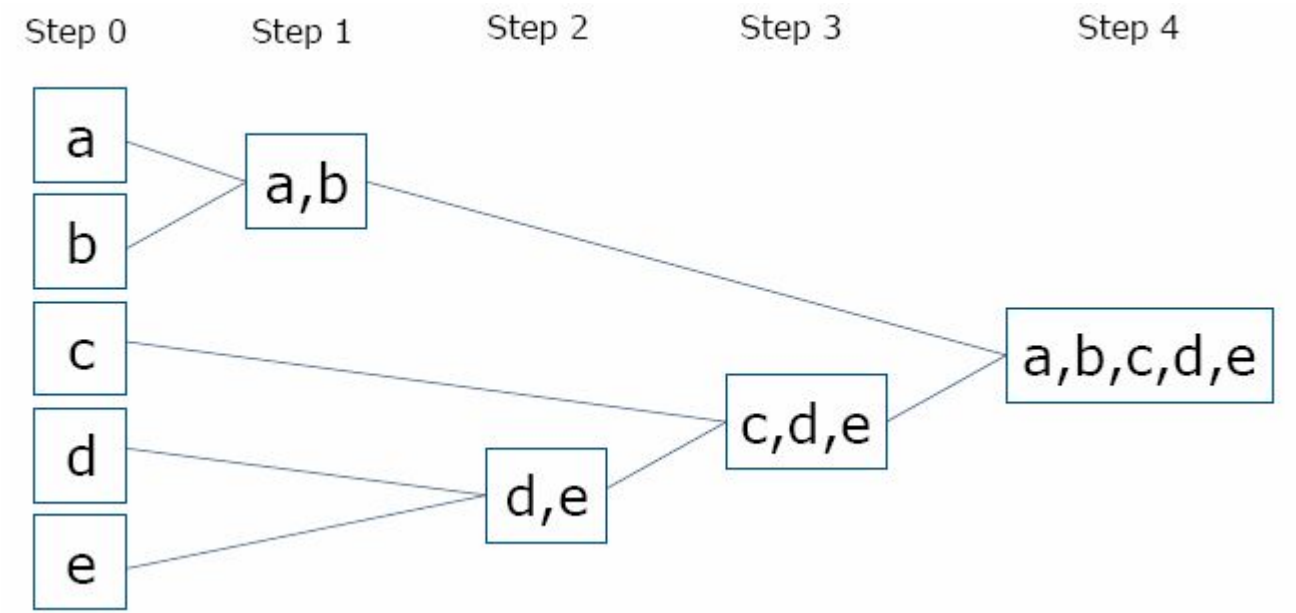
1. Métodos de Aprendizaje No Supervisado - Descripción
Clustering - PAM

- **Particionado Sobre los Medoides (PAM) o k-Medoids**
  - Objetos de un cluster cuya disimilaridad media al resto de objetos del cluster es mínima
- Trabaja, como el K-Means, con particiones, dividiendo el conjunto de datos en grupos
  - Ambos intentan minimizar la distancia entre puntos que se añadirían a un grupo y otro punto designado como el centro de ese grupo
  - k-medoids escoge datapoints como centros y trabaja con una métrica arbitraria de distancias entre datapoints
  - Es **más robusto** ante el ruido y a partes aisladas que k-means porque minimiza una suma de disimilaridades (entre pares de puntos) en vez de una suma de distancias euclidianas cuadradas.
- Un **medoid** puede ser definido como el objeto de un grupo cuya disimilaridad media a todos los objetos en el grupo es mínima
  - Es el punto ubicado más hacia el centro en todo el grupo.

1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Clúster jerárquico

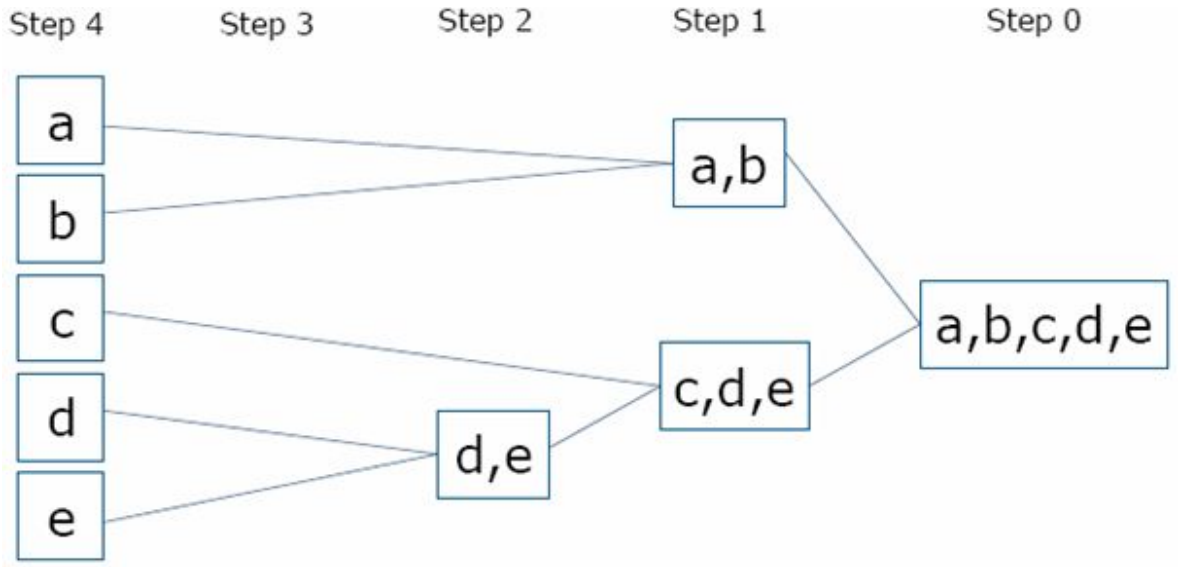
- Supón que tienes 5 elementos {a,b,c,d,e}
  - Inicialmente, consideramos cada uno, por sí mismo, un cluster
  - Entonces, en cada paso, tomamos los clústeres más similares entre sí, y los agrupamos en un nuevo cluster



1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Clúster jerárquico (II)

- Supón que tienes 5 elementos {a,b,c,d,e}
  - Inicialmente, consideramos un único cluster con todos los elementos dentro
  - Entonces, en cada paso, partimos un cluster para mejorar la distancia intra-cluster, hasta que todos los elementos son clústeres interdependientes



1. Métodos de Aprendizaje No Supervisado - Descripción

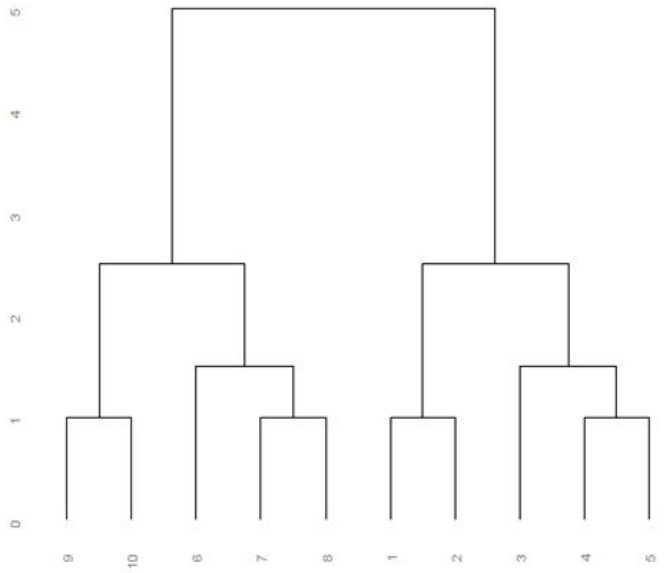
Clustering - Clúster jerárquico (III)

- Esos son los dos enfoques de clustering jerárquico, denominados
  - **Aglomerativo** ➜ comenzamos con clústeres individuales, que vamos uniendo entre sí
  - **Divisivo** ➜ comenzamos con un sólo cluster que vamos dividiendo hasta que todos los elementos pertenecen a clústeres independientes
- **Fortalezas**
  - No necesita que se establezca el número deseado de clústeres
  - Se pueden obtener divisiones en cualquier número de clústeres “cortando” el dendograma en el nivel apropiado
  - Pueden corresponderse con taxonomías reales
  - Utilizan una matriz de distancias o proximidad, para unir o dividir clústeres según ésta

1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Clúster jerárquico (IV)

- El algoritmo proporciona una jerarquía de particiones que se presentan en forma de árboles (dendogramas).
- La visión del árbol sugiere el número de grupos a retener.
- Cada corte del árbol da una partición. Cuanto más arriba se corte el árbol se obtendrá un menor número de clases y clases menos homogéneas.
- Ejemplo de ɔación de números del con distancia euclídea



1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Clúster jerárquico (V)

- **Divisiones**
  - **(a) Ligadura simple o vecino más cercano**: Distancia mínima. Encuentra los objetos separados por la distancia más corta y los agrupa. Objetos más similares. Puede formar grupos formados por individuos demasiado heterogéneos.
  - **(b) Ligadura completa o vecino más lejano**: Distancia máxima. Encuentra los objetos separados por la distancia mayor y los agrupa. Objetos menos similares. Evita el problema anterior.
  - **(c) Ligadura promedio**: Comienza igual que los anteriores, pero el criterio de unión es la distancia media de todos los individuos de un grupo con todos los individuos de otro. Semejanza media entre objetos.
  - **(d) Método del centroide**: La distancia entre dos grupos es la distancia entre sus centroides, que son los valores medios de las observaciones en el valor teórico del grupo. Cada vez que se forman los grupos se calculan de nuevo los centroides.
  - **(e) Métodos basados en varianza en vez de en distancias (Ward)**: Descomponen la varianza o inercia en inercia intra clase (dentro de la clase) e inercia inter clase (entre diferentes clases).

1. Métodos de Aprendizaje No Supervisado - Descripción

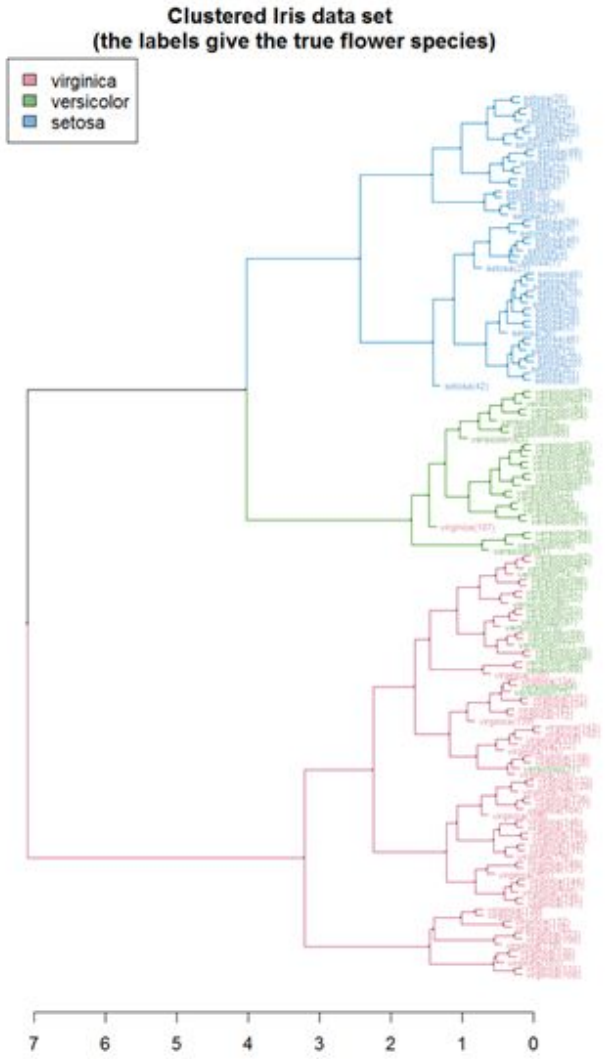
Clustering - Clúster jerárquico (VI)

- El criterio de Ward y k-medias sesgados hacia clusters con el mismo número de observaciones.
- El criterio de ligadura promedio está sesgado hacia clusters con la misma varianza.
- Menos sesgados son los métodos de ligadura simple y estimación de densidad (como k vecinos más próximos).
- Recomendación (Milligan (1981)):
  - Mejores resultados suelen provenir de los métodos de la ligadura promedio y el criterio de Ward
  - Peores los obtenidos con ligadura simple.

1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Clúster jerárquico (VII)

- Con el clustering jerárquico podemos obtener tantas divisiones como elementos haya
- ¿Cuál escogemos?



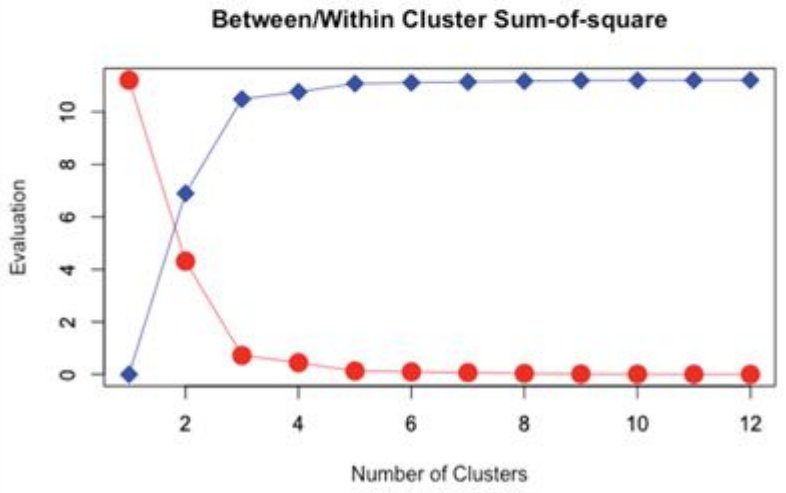
1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Clúster jerárquico (VIII)

- **Within-cluster sum of squares** (WSS)
  - Distancia de los elementos de un cluster a su centroide
- **Between-cluster sum of squares** (BSS)
  - Distancia entre centroides de clusters
- Dibujar **WSS** y **BSS** y buscar el punto con cambio significativo

$$\mathrm{WSS}(C) \sum_{i=1}^{k} \sum_{x_j \in C_i} d(x_j, \mu_i)^2$$

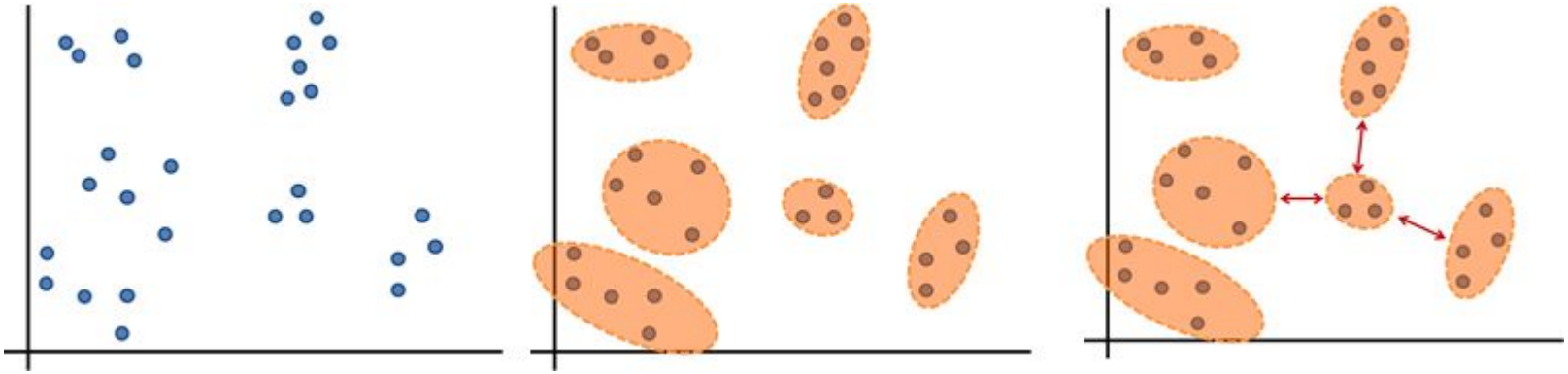
$$\mathrm{BSS}(C) \sum_{i=1}^{k} |C_i| d(\mu, \mu_i)^2$$



1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Clúster jerárquico (IX)

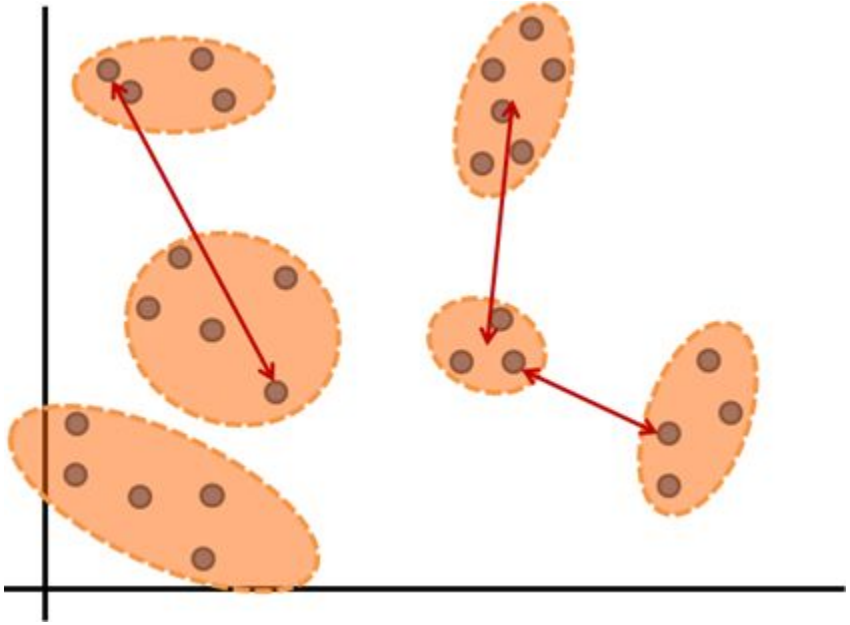
- *Funcionamiento*: si comenzamos con N clusters
  - Tendremos una matriz de distancias entre puntos
  - Empezaremos a agrupar
  - Y a actualizar esa matriz de distancias
  - Y en algún punto, esa matriz de distancias contendrá distancias entre clusters,
  - ¿Cómo la calculamos?



1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Clúster jerárquico (X)

- Medidas típicas:
  - Mínima distancia entre un elemento de un cluster y otro
  - Máxima distancia entre un elemento de un cluster y otro
  - Distancia promedio entre los elementos de los clusters
  - Distancia entre los centroides de los clusters
  - ...



1. Métodos de Aprendizaje No Supervisado - Descripción

Clustering - Representación de clústeres

- **Para datos numéricos**
  - Podemos identificar un cluster por su centroide (Punto promedio)
  - De una manera alternativa, por su envolvente convexa
- **Para datos no numéricos**
  - Podemos utilizar cualquier distancia
  - No podemos establecer un centroide ➜ clustoide
    - Es una instancia que se toma como representante del cluster
    - Puede ser el punto que minimiza la suma de las distancias con los otros puntos del cluster
      - O minimiza la distancia máxima a otro elemento
      - O minimiza la suma al cuadrado de las distancias con otros elementos
      - ...

1. Métodos de Aprendizaje No Supervisado - Descripción

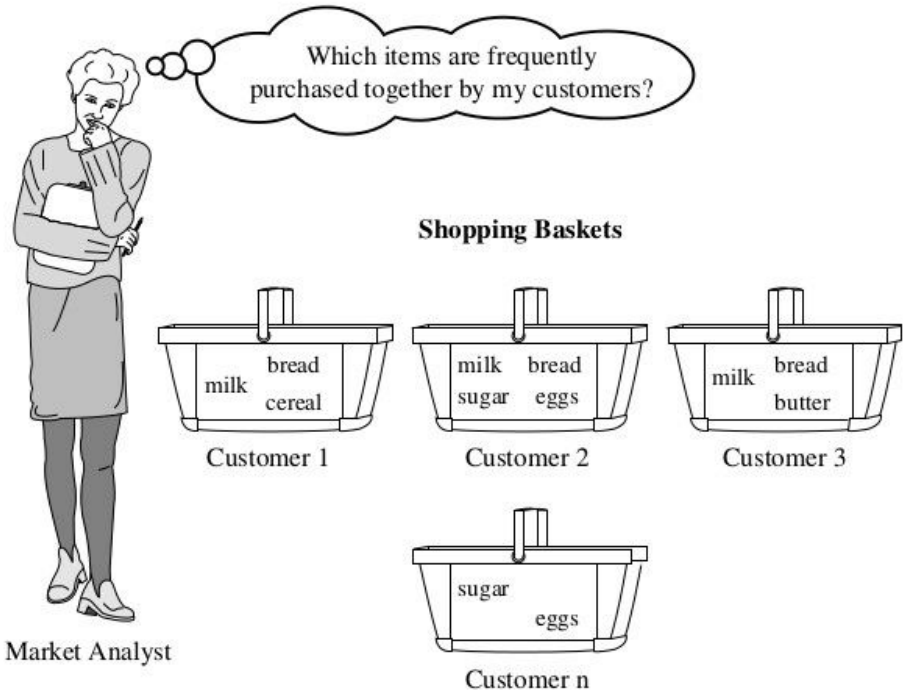
Clustering - Recomendaciones finales

- Procedimientos jerárquicos más rápidos y llevan menos tiempo de cálculo
  - Combinaciones indeseables iniciales pueden perdurar a lo largo del análisis.
  - Necesitan almacenar gran cantidad de datos.
- Procedimientos no jerárquicos dependen de las semillas o puntos de partida
  - Menos susceptibles a outliers (en los jerárquicos pueden formar un grupo por sí mismos más fácilmente).
- Una vez que se combinan 2 clusters, la decisión no es reversible
- Problemas ante:
  - Presencia de outliers, y datos con ruido
  - Clusters con tamaños muy diferentes
  - Partición de clusters que agrupan muchos elementos

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación

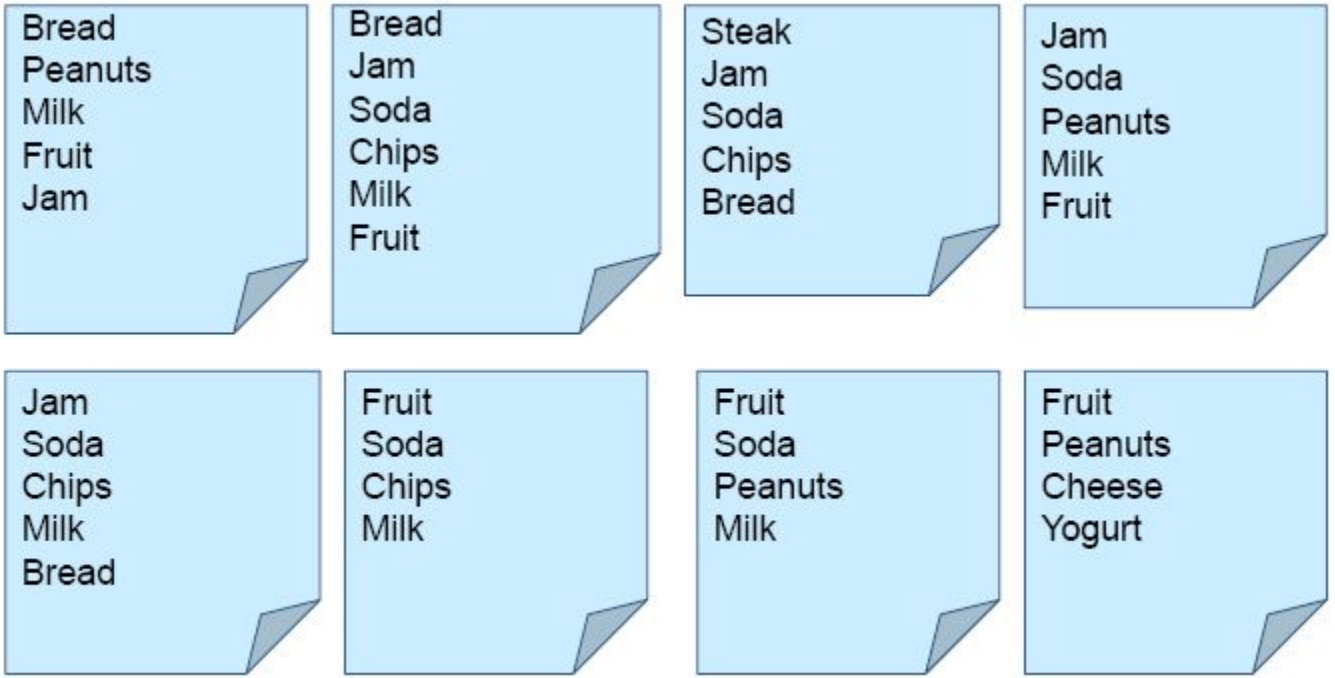
- Tratan de encontrar patrones frecuentes, asociaciones, correlaciones dentro de conjuntos de elementos u objetos en bases de datos
- Aplicaciones
  - Análisis de la cesta de la compra
  - Marketing cruzado
  - Diseño de catálogos
  - ...



1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación (II)

- **Cesta de la compra**
  - Encontrar parejas de productos que suelen aparecer juntos en la cesta



1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Extracción de reglas

- Dado un conjunto de transacciones, encontrar reglas que **describen** la **aparición** de un **elemento basándose** en la **aparición** de **otros**
- Ejemplos
  - {Bread} → {Milk}
  - {Soda} → {Chips}
  - {Bread} → {Jam}
  - ...

TID	Items
1	Bread, Peanuts, Milk, Fruit, Jam
2	Bread, Jam, Soda, Chips, Milk, Fruit
3	Steak, Jam, Soda, Chips, Bread
4	Jam, Soda, Peanuts, Milk, Fruit
5	Jam, Soda, Chips, Milk, Bread
6	Fruit, Soda, Chips, Milk
7	Fruit, Soda, Peanuts, Milk
8	Fruit, Peanuts, Cheese, Yogurt

- La implicancia indica **"*co-ocurrencia*" no causalidad**

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Definiciones

- **Itemset**
  - Conjunto de uno o más ítems
    - {milk, bread, jam}
  - K-itemset
    - Un itemset con exactamente k elementos
- **Soporte**
  - Porcentaje de tra contienen un itemset
    - Soporte({milk, bread}) = 3
    - Soporte({soda, chips}) = 3
- **Itemset frecuente**
  - Un itemset cuyo soporte es superior a determinado umbral

TID	Items
1	Bread, Peanuts, Milk, Fruit, Jam
2	Bread, Jam, Soda, Chips, Milk, Fruit
3	Steak, Jam, Soda, Chips, Bread
4	Jam, Soda, Peanuts, Milk, Fruit
5	Jam, Soda, Chips, Milk, Bread
6	Fruit, Soda, Chips, Milk
7	Fruit, Soda, Peanuts, Milk
8	Fruit, Peanuts, Cheese, Yogurt

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Métricas

- Implicación de la forma X→ Y, siendo X e Y itemsets
  - *{bread}* → *{milk}*
- Evaluación de la bondad de una regla
  - **Soporte**: porcentaje de transacciones que contienen ambos itemsets X e Y
  - **Confianza**: mide con qué frecuencia elementos de Y aparecen junto a X en una transacción

TID	Items
1	Bread, Peanuts, Milk, Fruit, Jam
2	Bread, Jam, Soda, Chips, Milk, Fruit
3	Steak, Jam, Soda, Chips, Bread
4	Jam, Soda, Peanuts, Milk, Fruit
5	Jam, Soda, Chips, Milk, Bread
6	Fruit, Soda, Chips, Milk
7	Fruit, Soda, Peanuts, Milk
8	Fruit, Peanuts, Cheese, Yogurt

$$s = \frac{\sigma(\{\text{Bread, Milk}\})}{\#\text{ of transactions}} = 0.38$$

$$c = \frac{\sigma(\{\text{Bread, Milk}\})}{\sigma(\{\text{Bread}\})} = 0.75$$

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Métricas (II)

- Observaciones
  - Todas las reglas se originaron en el itemset: {Milk, Diaper, Beer}
  - Las reglas que se originan en el mismo itemset tienen el mismo soporte pero pueden tener distinta confianza ¿Por qué?

TID	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

```
{Milk,Diaper} → {Beer} (s=0.4, c=0.67)
{Milk,Beer} → {Diaper} (s=0.4, c=1.0)
{Diaper,Beer} → {Milk} (s=0.4, c=0.67)
{Beer} → {Milk,Diaper} (s=0.4, c=0.67)
{Diaper} → {Milk,Beer} (s=0.4, c=0.5)
{Milk} → {Diaper,Beer} (s=0.4, c=0.5)
```

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Métricas (III)

- Dado un **conjunto de transacciones**:
  - Encontrar reglas que superen unos umbrales establecidos de soporte y confianza

TID	Items
1	Bread, Peanuts, Milk, Fruit, Jam
2	Bread, Jam, Soda, Chips, Milk, Fruit
3	Steak, Jam, Soda, Chips, Bread
4	Jam, Soda, Peanuts, Milk, Fruit
5	Jam, Soda, Chips, Milk, Bread
6	Fruit, Soda, Chips, Milk
7	Fruit, Soda, Peanuts, Milk
8	Fruit, Peanuts, Cheese, Yogurt

{Bread, Jam} ⇒ {Milk} s=0.4 c=0.75
{Milk, Jam} ⇒ {Bread} s=0.4 c=0.75
{Bread} ⇒ {Milk, Jam} s=0.4 c=0.75
{Jam} ⇒ {Bread, Milk} s=0.4 c=0.6
{Milk} ⇒ {Bread, Jam} s=0.4 c=0.5

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Objetivo

- **(1) Encontrar conceptos relacionados**
  - Supongamos que las palabras son los ítems y los “documentos” las canastas.
- **(2) Plagio**
  - En este caso los ítems son las documentos y las canastas las oraciones.
  - Donde un “*item/documento*” está en una “*canasta/oración*” si la oración pertenece al documento.
  - Una o dos oraciones en común en distintos documentos son un buen indicador de plagio.

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Enfoque de dos pasos

- Se suelen basar en variaciones de un enfoque en dos pasos:
  - ***Paso 1*: Generar aquellos itemsets frecuentes**
    - Con soporte mayor que un determinado umbral
    - Computacionalmente caro
  - ***Paso 2:* Generación de reglas**
    - Generar reglas con alta confianza haciendo uso de los itemsets frecuentes obtenidos en el paso anterior
    - Cada regla es una partición binaria de un itemset frecuente
    - Abajo tenemos todas las particiones binarias del itemset {Bread, Jam, Milk}

{Bread, Jam} ⇒ {Milk} s=0.4 c=0.75
{Milk, Jam} ⇒ {Bread} s=0.4 c=0.75
{Bread} ⇒ {Milk, Jam} s=0.4 c=0.75
{Jam} ⇒ {Bread, Milk} s=0.4 c=0.6
{Milk} ⇒ {Bread, Jam} s=0.4 c=0.5

1. Métodos de Aprendizaje No Supervisado - Descripción

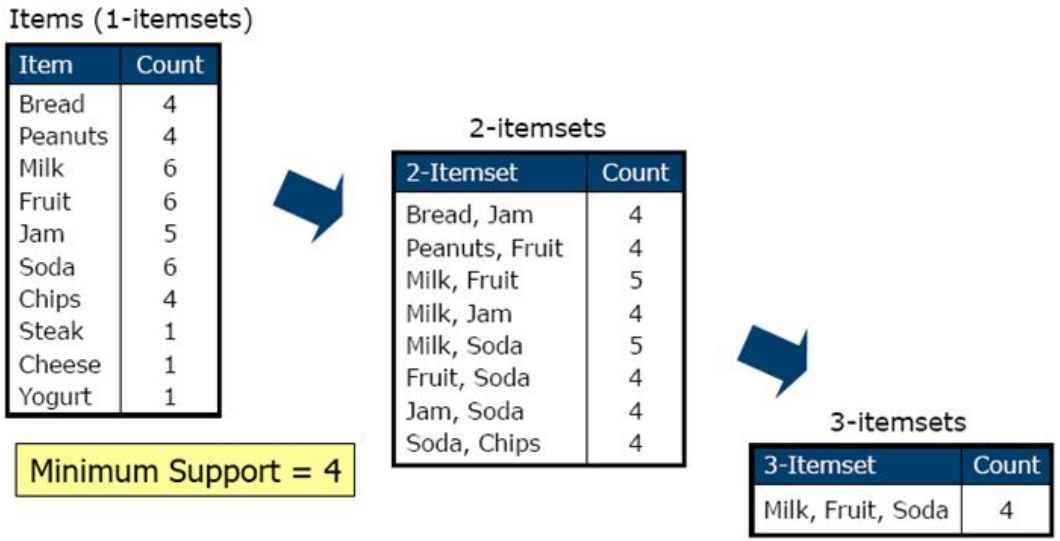
Reglas de asociación - Complejidad computacional

- Hay muchos itemsets candidatos a explorar (2^N)
  - Para 3 productos {ABC}
    - {A}, {B}, {C}, {AB}, {AC}, {BC}, {ABC}
  - Para 4 productos {ABCD}
    - {A}, {B}, {C}, {D}, {AB}, {AC}, {AD}, {BC}, {BD}, {CD}, {ABC}, {ABD}, {ACD}, {BCD}, {ABCD}
- Es intratable cuando el número de elementos crece
  - Para 25 productos ➜ 33554432
  - Para 100 productos ➜ 1.2676506e+30

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Algoritmo Apriori

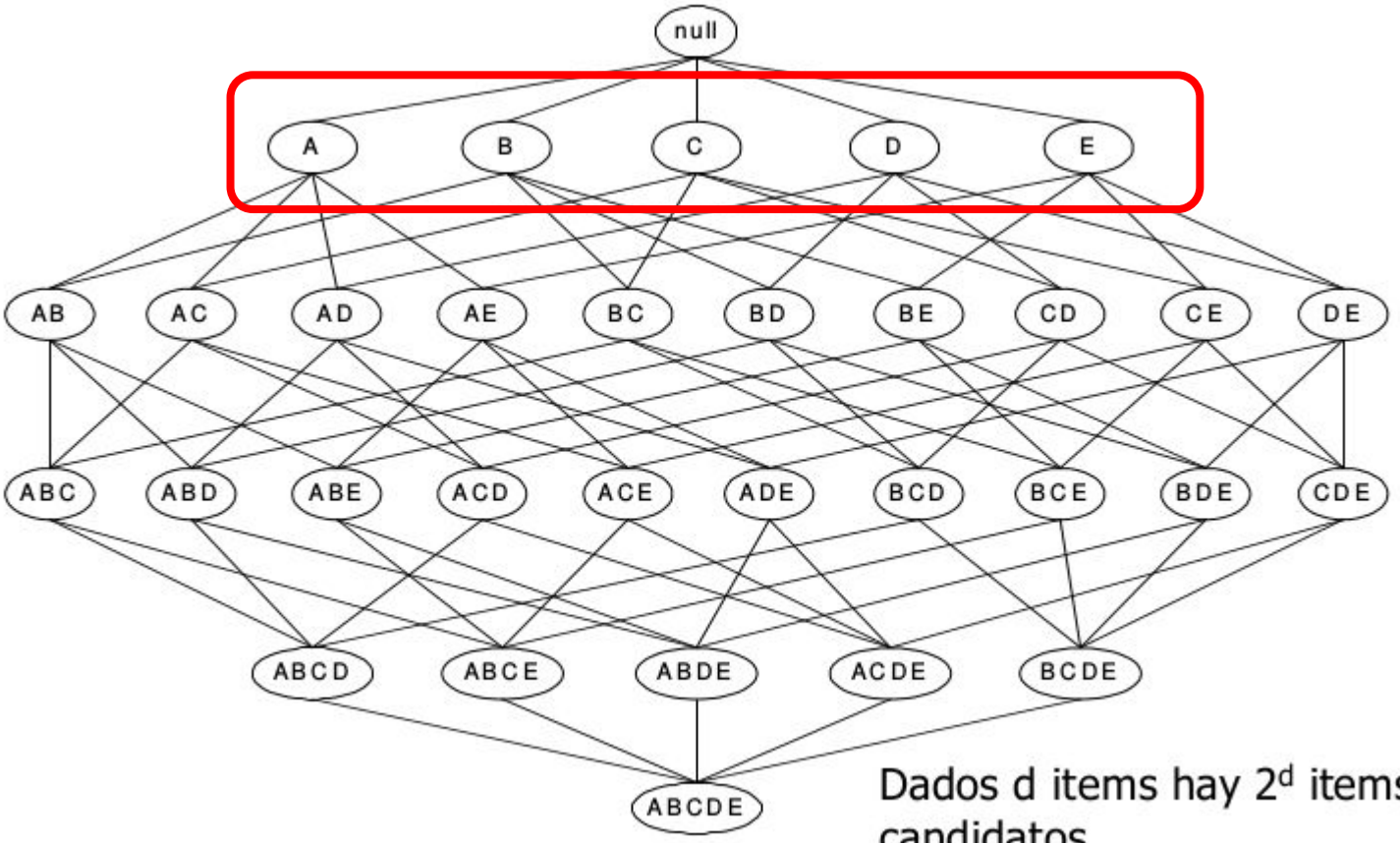
- Fue propuesto por Agrawal y Srikant en 1994 para hacer minería de datos sobre itemsets (elementos) frecuentes con reglas de asociación binarias (como los ejemplos de la canasta).
- Comenzar con itemsets de tamaño k=1, ir incrementando el valor de k de 1 en 1, descartando aquellos itemsets que no cumplan un soporte mínimo



1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Algoritmo Apriori (II)

- ***Paso 1*: Generación de itemsets frecuentes**



1. Métodos de Aprendizaje No Supervisado - Descripción
Reglas de asociación - Algoritmo Apriori (III)

- ***Paso 1*: Generación de itemsets frecuentes**
  - Estrategias para la generación de itemsets
    - **Reducir el número de candidatos (M)**
      - Búsqueda completa: M = 2d
      - Utilice técnicas de poda para reducir M
    - **Reducir el número de transacciones (N)**
      - Reducir el tamaño de N como el incremento del tamaño de los itemsets
      - Esto es utilizado algoritmos como Direct Hashing and Pruning (DHP)
    - **Reducir el número de comparaciones (NM)**
      - Utilice las estructuras de datos eficientes para almacenar los candidatos o transacciones
      - No hay necesidad de comparar cada candidato contra cada transacción

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Algoritmo Apriori (IV)

- ***Paso 2*: Generación de reglas**
  - Dado un conjunto de transacciones T, el objetivo del descubrimiento de reglas de asociación es encontrar todas las reglas que cumplen:
    - **support** ≥ **minsup** threshold
    - **confidence** ≥ **minconf** threshold
  - Aproximación de fuerza bruta:
    - Listar todas las posibles reglas de asociación
    - Calcular el soporte y la confianza para cada una
    - Eliminar las que no satisfacen los umbrales predefinidos

⇒ Computacionalmente Prohibitivo!

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Algoritmo Apriori (V)

- ¿Por qué es prohibitivo?
  - Supongamos que tenemos *frecuent itemset* de 100 items:

$$\{a\,1\,,\,a\,2\,,\,\ldots\,,\,a\,100\,\}$$

- Vamos a tener los $\binom{100}{1}=100$ 1-itemset frecuentes
- Vamos a tener los $\binom{100}{2}=4950$ 2-itemset frecuentes

$$\binom{100}{1}+\binom{100}{2}+\cdots+\binom{100}{100}=2^{100}-1\approx 1.27\times 10^{30}.$$

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Algoritmo Apriori (VI)

- Dado un itemset frecuente, encontrar reglas que tienen una mínima confianza
  - Para 3 Productos {ABC}
    - A → BC, B→AC, C→ AB, AB→ C, AC→ B, BC→ A
  - Es intratable cuando el número de elementos crece
    - Para 25 productos → 33554430 (2 menos que antes)
- Al igual que antes, el algoritmo Apriori deja de explorar alternativas que presentan baja confianza

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Algoritmo Apriori (VII)

- **Principio Apriori**
  - Si un itemset es frecuente, entonces todos sus subsets deben además ser frecuentes.
  - El Principio Apriori se sostiene debido a las siguiente propiedades de la medida de support:

$$\forall X, Y : (X \subseteq Y) \Rightarrow s(X) \geq s(Y)$$

  - El **support** de un itemset nunca excede el **support** de sus subsets
  - Esto es conocido como la propiedad de *anti-monotonía* del **support**
- **Anti-monotonía**
  - Si un itemset X no satisface el umbral de min_support entonces X no es frecuente.
  - Es decir:

$$S(X) < min\_support$$

  - Si agrego X2 al itemset X (X U X2) entonces el resultado del itemset no puede ser más frecuente que X
  - X U X2 es no frecuente, por lo tanto:

$$S(X \cup X2) < min\_support$$

1. Métodos de Aprendizaje No Supervisado - Descripción

Reglas de asociación - Algoritmo Apriori (VIII)

- Si el soporte mínimo es muy grande podemos perder itemsets que incluyan elementos de interés, pero poco frecuentes (productos caros)
- Si el soporte mínimo es muy bajo, es computacionalmente muy difícil de calcular (gran número de itemsets)
- Heurística
  - Comenzar con un soporte alto, e ir reduciendo hasta obtener un número de reglas adecuado
  - Hasta que sea computacionalmente factible ejecutar el algoritmo

1. Métodos de Aprendizaje No Supervisado - Descripción

Reducción de la dimensionalidad - Motivación

- ¿**Qué atributos** deben de **usarse** para hacer modelos predictivos?
  - En términos generales, responder a esto requiere de un profundo conocimiento del problema, y de los datos.
- Nos encontramos con dos respuestas para esta pregunta
  - Las técnicas de selección de atributos buscan seleccionar aquellos atributos que son más relevantes para un determinado problema.
  - Las técnicas de reducción de la dimensionalidad buscan crear nuevos atributos mediante combinaciones de varios ya existentes.

1. Métodos de Aprendizaje No Supervisado - Descripción

Reducción de la dimensionalidad - Motivación (II)

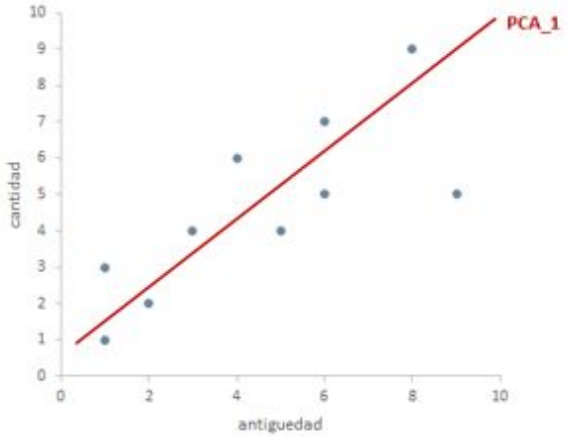
- Un problema central en el análisis de datos multivariantes es la reducción de la dimensionalidad
- Si es posible describir con precisión los valores de p variables por un **pequeño subconjunto *r* < *p* de ellas**, se habrá reducido la dimensión del problema a costa de una pequeña pérdida de información
- El Análisis de Componentes Principales (ACP) tiene ese objetivo
  - Dadas n observaciones de p variables, se analiza si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales

1. Métodos de Aprendizaje No Supervisado - Descripción

Reducción de la dimensionalidad - Motivación (III)

- Por ejemplo, con variables de alta dependencia es frecuente que un pequeño número de nuevas variables (menos del 20% de las originales) expliquen la mayor parte (el 80% de la variabilidad original)

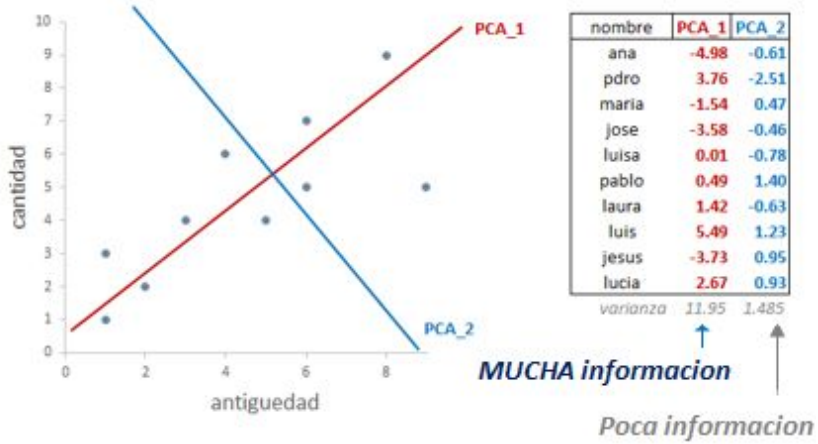
PCA 1



nombre	PCA_1
ana	-4.98
pdro	3.76
maria	-1.54
jose	-3.58
luisa	0.01
pablo	0.49
laura	1.42
luis	5.49
jesus	-3.73
lucia	2.67

En este ejemplo el PCA_1 tiene el **89%** de la varianza

PCA 1 + PCA 2



nombre	PCA_1	PCA_2
ana	-4.98	-0.61
pdro	3.76	-2.51
maria	-1.54	0.47
jose	-3.58	-0.46
luisa	0.01	-0.78
pablo	0.49	1.40
laura	1.42	-0.63
luis	5.49	1.23
jesus	-3.73	0.95
lucia	2.67	0.93
varianza	*11.95*	*1.485*

En este ejemplo el PCA_2 tiene el **11%** de la varianza (1.485/(11.95+1.48))

Fuente: https://3.bp.blogspot.com/-YfV5P_qZ-5M/V3iF0Fm0N6I/AAAAAAAAD9g/XT5r1mNt8oE54atyJ4048YDftSPiNjG6gCLcB/s1600/pca2.png

1. Métodos de Aprendizaje No Supervisado - Descripción

Reducción de la dimensionalidad - Historia

- Su utilidad es doble:
  - Permite **representar óptimamente** en un espacio de dimensión más pequeño (y óptimo por lo tanto) las observaciones de algo (compras, ventas, transacciones, etc.) que esté expresado en un espacio p-dimensional (donde p es el número de características o variables que tiene). Naturalmente, sin que se produzca una pérdida de información.
    - En este sentido, los componentes principales es el primer paso para identificar las posibles **variables latentes**, o no observadas que generan los datos
  - Permite **transformar las variables originales**, en general correladas, en nuevas variables interreladas, facilitando la interpretación de datos
- La técnica de ACP es debida a Hotelling (1933)
- Sin embargo, sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados introducidos por K. Pearson (1901)

1. Métodos de Aprendizaje No Supervisado - Descripción

Reducción de la dimensionalidad - Historia (II)

Karl Pearson (1857-1936) *Científico británico. Inventor del contraste que lleva su nombre y uno de los fundadores de la Estadística en el siglo* XIX. *Sus trabajos sobre ajustes ortogonales precedieron el análisis de componentes principales. Fue Catedrático de matemáticas y después de Eugenesia en la Universidad de Londres. Fundador con Weldon, y con el apoyo económico de Galton, de la prestigiosa revista de estadística* Biometrika.

1. Métodos de Aprendizaje No Supervisado - Descripción

Reducción de la dimensionalidad - ACP

- ACP = Análisis de Componentes Principales
- A menudo tenemos datos sobre un colectivo con muchos atributos continuos entre los que hay cierta relación lineal (correlación).
  - Ejemplo: cantidades compradas de ciertos productos
- En el caso más extremo alguna de las variables podría ser una combinación lineal exacta de otra u otras
- Así, el objetivo es **transformar un conjunto de variables** (originales) en un **nuevo conjunto de variables** (componentes principales), **incorrelacionadas entre sí**

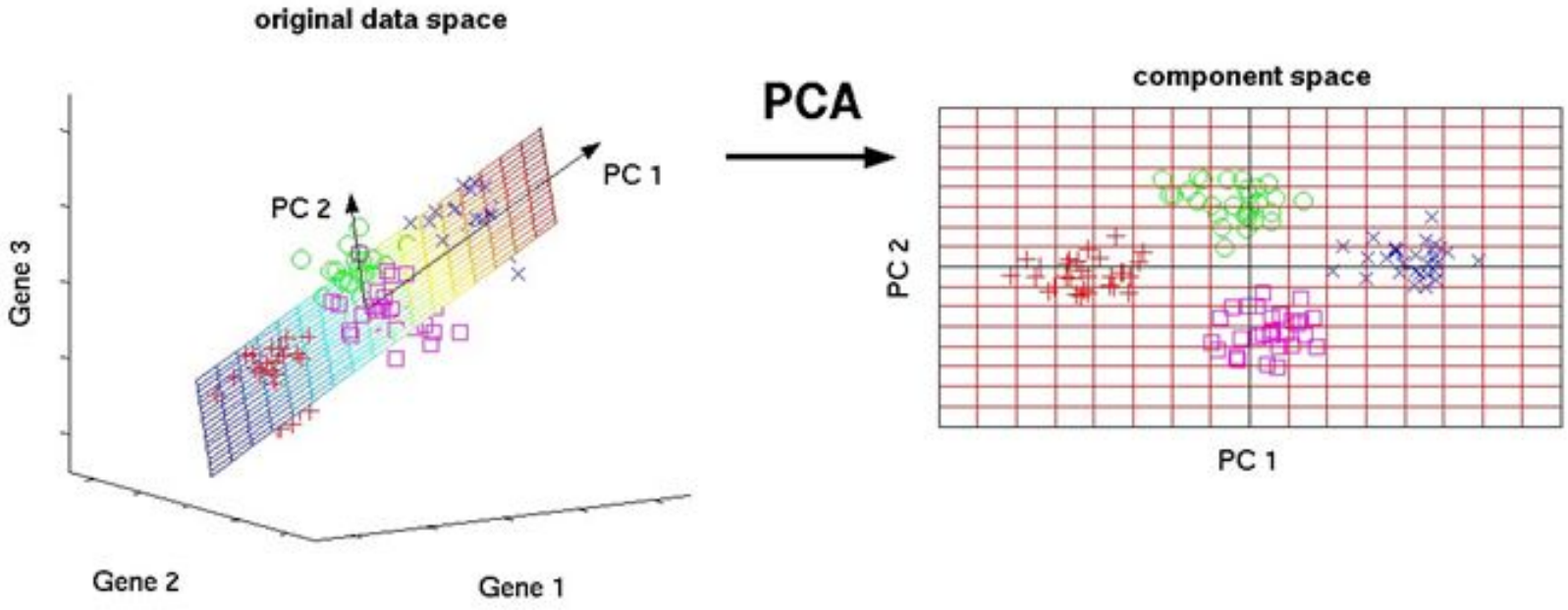
1. Métodos de Aprendizaje No Supervisado - Descripción
Reducción de la dimensionalidad - ACP (II)

- ¿No sería más útil tomar un subconjunto de las variables originales —o un número reducido de variables transformadas de las originales que describiera el colectivo sin gran pérdida de información?
  - A estas nuevas variables (combinación lineal de las originales) se les denomina componentes principales.
  - Se trata de tratar de explicar la misma información contenida en los datos originales con un menor número de variables buscadas de manera óptima
  - No hay ningún modelo o supuesto implícito.
  - Es importante tener en cuenta que las variables originales tienen que estar correlacionadas para que la reducción de la dimensión sea efectiva.

1. Métodos de Aprendizaje No Supervisado - Descripción

Reducción de la dimensionalidad - ACP (III)

- Cuestiones a resolver:
  - ¿Qué significa "*sin gran pérdida de información*"?
  - ¿Qué nuevas variables, distintas de las originales, estamos dispuestos a considerar?



Fuente: https://userscontent2.emaze.com/images/d1bbf2e6-271a-4500-9c0a-b3932515f1f1/09849cbfcd49ae092ecf83ebca4014d4.png

1. Métodos de Aprendizaje No Supervisado - Descripción

ACP - Análisis de interdependencias

- El interés viene de:
  - Para explicar fenómenos cuya información se cifra en **muchas variables** más o menos **correlacionadas**.
  - Reducir la **dimensión** del número de variables inicialmente consideradas en el análisis.
  - Las nuevas variables pueden **ordenarse** según la información que llevan
- Sólo con **datos cuantitativos** y no es necesario establecer jerarquías ni comprobar la normalidad

1. Métodos de Aprendizaje No Supervisado - Descripción
ACP - Análisis de interdependencias (II)

- Como medida de la cantidad de información incorporada en el componente se utiliza la varianza (б).
- Por tanto se ordenan según la varianza de mayor a menor
- Se trabaja con **variables tipificadas** o con **variables expresadas en desviaciones respecto a la media** para evitar problemas derivados de la escala
- **Utilidad** → si las variables que explican un fenómeno son muchas y están correlacionadas, es posible explicar el fenómeno con muy pocos componentes principales
  - Análisis de mercados
  - Determinación del proceso de compra
  - Caracterización de explotaciones
  - etc.

1. Métodos de Aprendizaje No Supervisado - Descripción
ACP - Ejemplo

- ***Ejemplo***: tenemos datos de las calificaciones obtenidas por estudiantes en:
  - Matemáticas
  - Inglés
  - Ciencias Naturales
- Calculamos la matriz de correlación y obtenemos:
  - ¿Qué nos dice?

$$R = \begin{pmatrix} 1.00 & 0.68 & 0.92 \\ 0.68 & 1.00 & 0.57 \\ 0.92 & 0.57 & 1.00 \end{pmatrix}$$

- Cada estudiante podría ser descrito por **dos variables**:
  - Una reflejando su aptitud/interés por las Matemáticas y Ciencias Naturales (¿la nota media en ambas?)
  - Otra reflejando su aptitud/interés por idiomas.
- **Razonamiento implícito**:
  - 2 variables presentan elevada correlación, lo que sugiere que la información que aportan es muy redundante.
  - Conocido el valor que toma una podríamos conocer con bastante aproximación el valor que toma la otra.

1. Métodos de Aprendizaje No Supervisado - Descripción

ACP - Ejemplo (II)

- ***Ejemplo***: Estudiar el beneficio y la dimensión de 9 explotaciones bovinas ecológicas
  - ¿Qué variables vamos a utilizar?
    - Beneficio: € por explotación
    - Dimensión: inversión € por explotación
  - *Primer paso*: ¿hay correlación entre ambas variables?
  - *Segundo paso*: eliminar el problema de la escala
  - *Tercer paso*: obtener los componentes principales

	variables originales	
explotación	inversión (€)	beneficios (€)
1	775.104	23.795
2	775.218	58.778
3	700.963	1.531
4	674.063	-12.756
5	631.003	14.729
6	537.744	9.059
7	489.155	12.541
8	448.465	13.495
9	445.853	-34.828

1. Métodos de Aprendizaje No Supervisado - Descripción

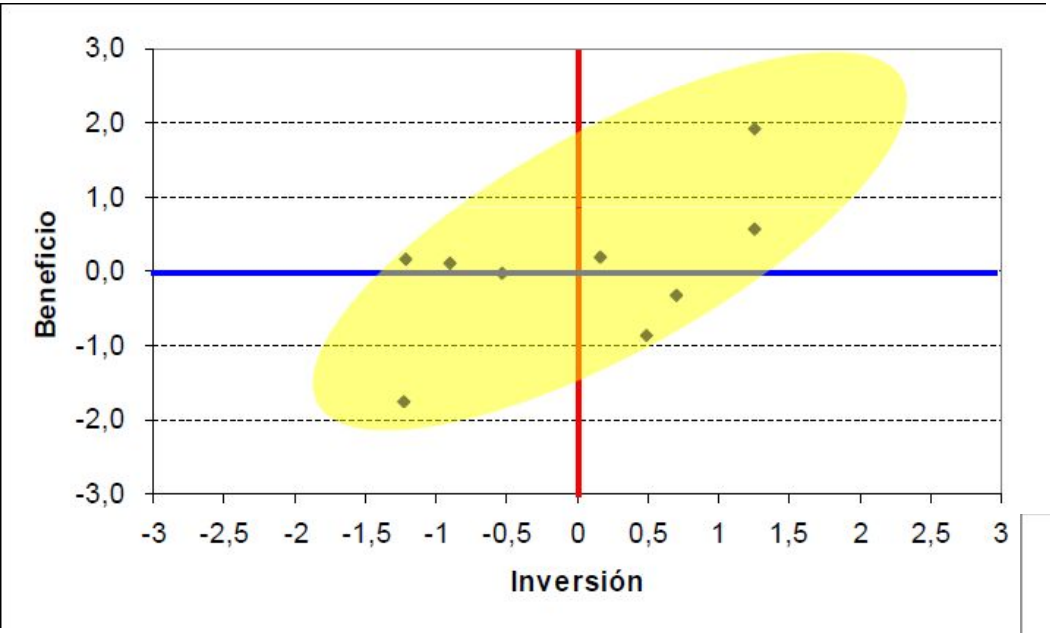
ACP - Ejemplo (III)

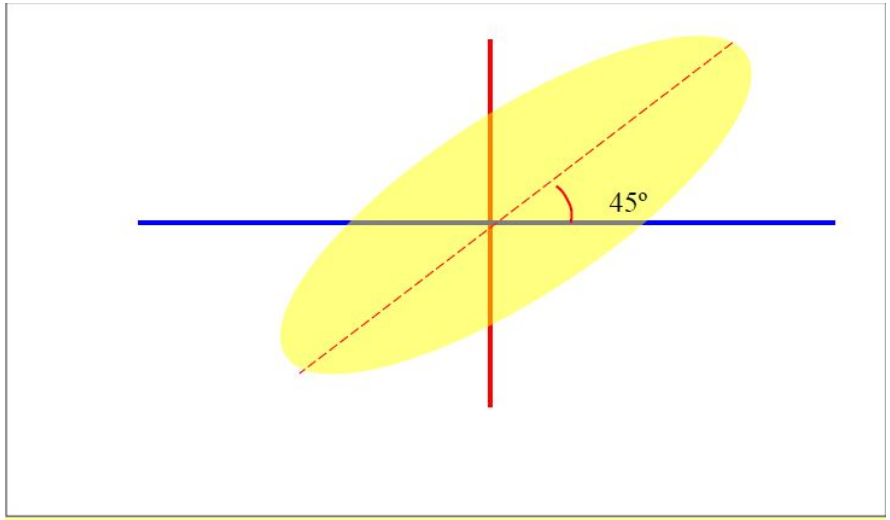
- *Segundo paso*: eliminar el problema de la escala
  - Tipificar las variables
  - La matriz de correlación es igual a la matriz de covarianzas
  - Σ бcomponentes principales = Σ бvariables = Σ variables tipificadas (2 en este caso)

	variables originales		variables tipificadas	
explotación	inversión (€)	beneficios (€)	inversión	beneficios
1	775.104	23.795	1,257	0,556
2	775.218	58.778	1,258	1,927
3	700.963	1.531	0,697	-0,316
4	674.063	-12.756	0,494	-0,875
5	631.003	14.729	0,169	0,201
6	537.744	9.059	-0,535	-0,020
7	489.155	12.541	-0,902	0,115
8	448.465	13.495	-1,209	0,152
9	445.853	-34.828	-1,229	-1,749

1. Métodos de Aprendizaje No Supervisado - Descripción

ACP - Ejemplo (IV)





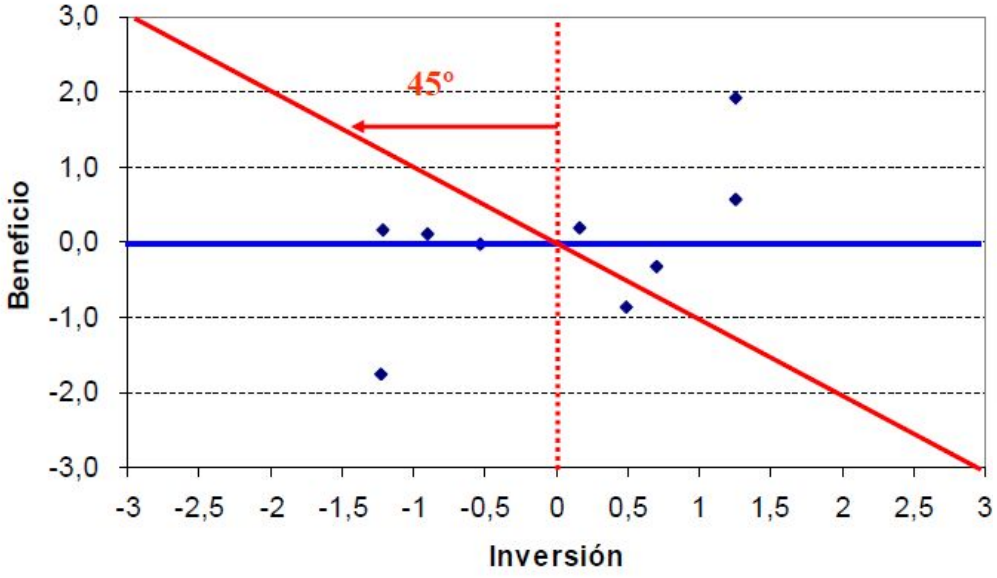
Elipse de concentración en una distribución normal bivariante

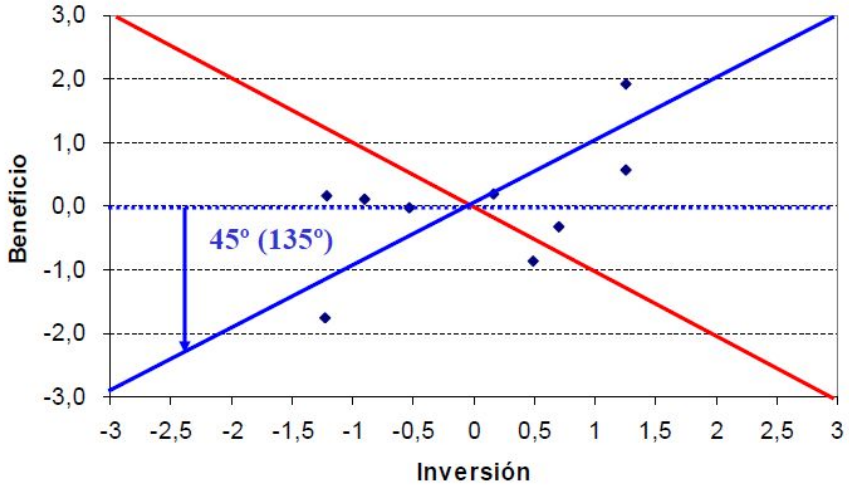
1. Métodos de Aprendizaje No Supervisado - Descripción
ACP - Ejemplo (V)

- *Tercer paso*: obtener componentes principales
  - Los coeficientes de los vectores son los coeficientes que hay que aplicar a las variables tipificadas para obtener los CP:
    - CP1 = u11* X1 + u12* X2
    - CP2 = u21* X1 + u22* X2
  - En el ejemplo
    - CP1 = 0,7071 * inversión + 0,7071 * beneficio
    - CP2 = 0,7071 * inversión – 0,7071 * beneficio
  - Son los senos y los cosenos del ángulo de rotación entre los ejes de los CP y los ejes de las variables tipificadas
    - Primer eje: cos 45º = 0,7071 sen 45º = 0,7071
    - Segundo eje: cos 135º = 0,7071 sen 135º = -0,7071

1. Métodos de Aprendizaje No Supervisado - Descripción
ACP - Ejemplo (VI)

- *Tercer paso*: obtener componentes principales





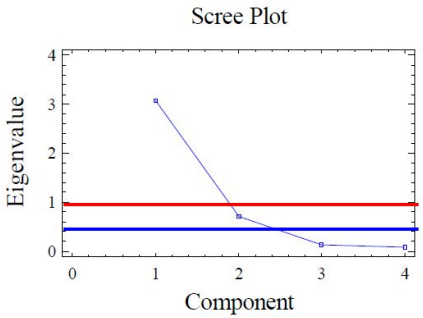
1. Métodos de Aprendizaje No Supervisado - Descripción
ACP - Ejemplo (VII)

- *Tercer paso*: obtener componentes principales
  - Ahora hay que determinar las cargas factoriales
    - Correlación de cada variables con cada CP
    - Coeficiente de correlación rhj entre el componente h y la variable j:
      - rhj = uhj * √λh
    - En nuestro caso (matriz factorial o matriz de componentes):
      - CP1 con inversión: 0,7071 * √1,54603 = 0,87821
      - CP1 con beneficio: 0,7071 * √1,54603 = 0,87821
      - CP2 con inversión: 0,7071 * √0,43397 = 0,47643
      - CP1 con beneficio: -0,7071 * √0,43397 = -0,47643

	variables originales		variables tipificadas		componentes principales	
explotación	inversión (€)	beneficios (€)	inversión	beneficios	CP1	CP2
1	775.104	23.795	1,257	0,556	1,283	-0,496
2	775.218	58.778	1,258	1,927	2,253	0,473
3	700.963	1.531	0,697	-0,316	0,270	-0,717
4	674.063	-12.756	0,494	-0,875	-0,270	-0,969
5	631.003	14.729	0,169	0,201	0,262	0,023
6	537.744	9.059	-0,535	-0,020	-0,393	0,364
7	489.155	12.541	-0,902	0,115	-0,556	0,720
8	448.465	13.495	-1,209	0,152	-0,747	0,963
9	445.853	-34.828	-1,229	-1,749	-2,100	-0,362

1. Métodos de Aprendizaje No Supervisado - Descripción
ACP - Ejemplo (VIII)

- *Tercer paso*: obtener componentes principales
  - **(1) Obtención de la primera componente**
    - Se busca que la primera componente tenga varianza máxima, considerando que la suma de los pesos (u1j) al cuadrado sea igual a la unidad.
    - La varianza a maximizar sería u´1Vu1 (siendo V la matriz de correlaciones).
    - Al resolver la ecuación (primera derivada de u1 e igualar a 0): $(V - \lambda I)u1 = 0$, se obtienen p raíces características y se toma la mayor (λ1).
    - Con la mayor λ1 se obtiene el vector asociado u1 que corresponde al vector característico asociado a la raíz característica mayor de la matriz V.
  - **(2) Obtención de los restantes componentes**
    - Se repite el mismo proceso aunque se impone la restricción de que cada nuevo vector sea ortogonal a los anteriores:
      - u`hu1 = u`hu2 = ... = u`huh-1 = 0
    - Por tanto, las p CP que se pueden calcular siempre son combinación lineal de las variables originales y los coeficientes de ponderación son los correspondientes vectores característicos asociados a la matriz V (matriz de covarianzas).
  - **(3) Número de componentes a retener**
    - Criterio de la media aritmética. Se seleccionan aquellas CP cuya raíz característica (λ) supere la media de las raíces características.
      - Si tenemos variables tipificadas, todas aquellas que superen el valor 1.
    - Contraste sobre raíces no retenidas. (suponer que las variables originales siguen una distribución normal)
    - El gráfico de sedimentación.



1. Métodos de Aprendizaje No Supervisado - Descripción
ACP - Ejemplo (IX)

- *Tercer paso*: obtener componentes principales
  - **(4) Retención de variables**
    - Si alguna de las variables presenta correlaciones muy débiles con los componentes retenidos, lo ideal es eliminarla del estudio (no está representada en ningún CP).
    - Si es una variable importante a juicio del investigador, habría que retener CP hasta que dicha variable estuviese representada
  - **(5) Interpretación de los componentes**
    - El investigador debe conocer el tema en profundidad para interpretar qué parte de la variabilidad explica el CP.
    - Para que un CP sea fácilmente interpretable:
      - Los coeficientes deben ser próximos a 1
      - Una variable debe tener un coeficiente elevado con sólo un CP
      - No deben existir CP con coeficientes similares

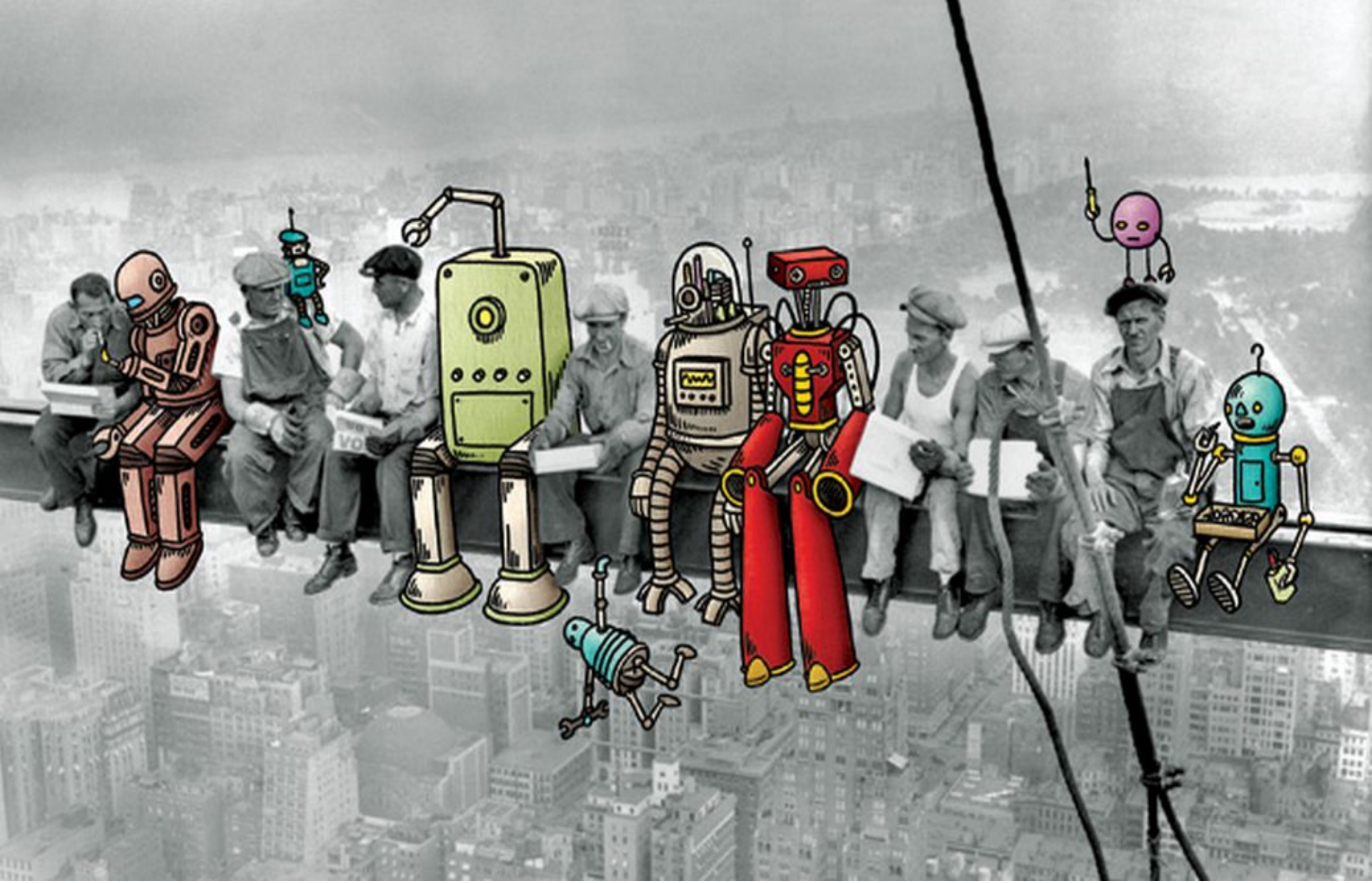
1. Métodos de Aprendizaje No Supervisado - Descripción
ACP - Elementos

- ¿Cómo se obtienen las componentes principales?
  - Buscando **combinaciones lineales** de las variables originales que **sean incorreladas entre sí** y que tengan la varianza máxima entre todas las posibles combinaciones lineales de las variables originales incorreladas con las nuevas variables o componentes principales
    - ROTACIÓN VARIMAX
- ¿Cuántas componentes principales necesitamos?
  - En realidad se calculan el **mismo número de componentes principales que de variables originales** Subselección:
    - Es una cuestión subjetiva: tantas como sean necesarias para obtener un porcentaje explicado de la información total “razonable”

1. Métodos de Aprendizaje No Supervisado - Descripción

ACP - Elementos (II)

- Emplear la matriz de correlaciones es equivalente a centrar y tipificar las variables y utilizar la matriz de covarianzas.
  - Se usa cuando las variables están medidas en escalas diferentes o cuando son heterogéneas y medidas en diferentes unidades.
  - Así participan de forma idéntica en el análisis.
  - Las componentes no son sino combinaciones lineales de las variables originales, las más correlacionadas con ellas.
  - Matemáticamente el problema se resuelve calculando los **valores y vectores propios** de la **matriz de covarianzas/correlaciones**
    - Las **componentes principales** surgen de los vectores propios



1. Métodos de Aprendizaje No Supervisado - Descripción

DP01 - Machine Learning (II)

DP - Data Proficiency - Data Analytics Journey