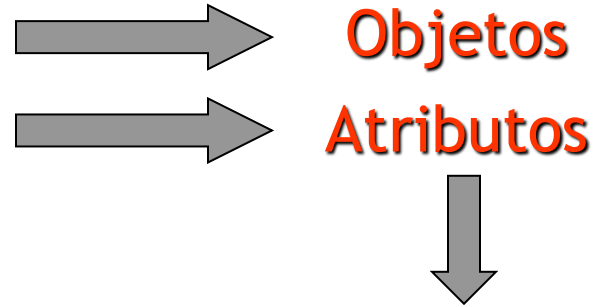


Laboratorio de Inteligencia Computacional

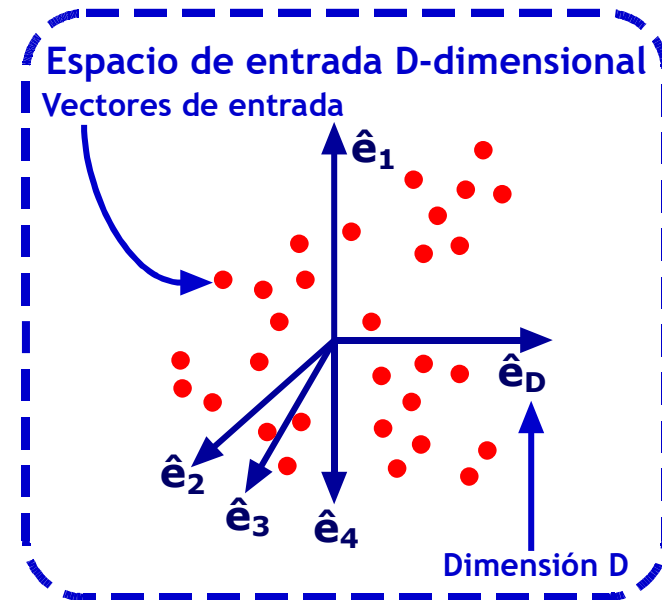
11 de Agosto de 2022

Cristián J. Figueroa

Clustering



Vectores en un Espacio de
Entrada Multidimensional



Corresponde a la organización de un
conjunto de patrones de entrada en grupos,
basado en una medida de similitud

Aplicaciones:

Segmentación de Imágenes y Texturas

Reconocimiento de Patrones

Data Mining y Text Mining

Marketing

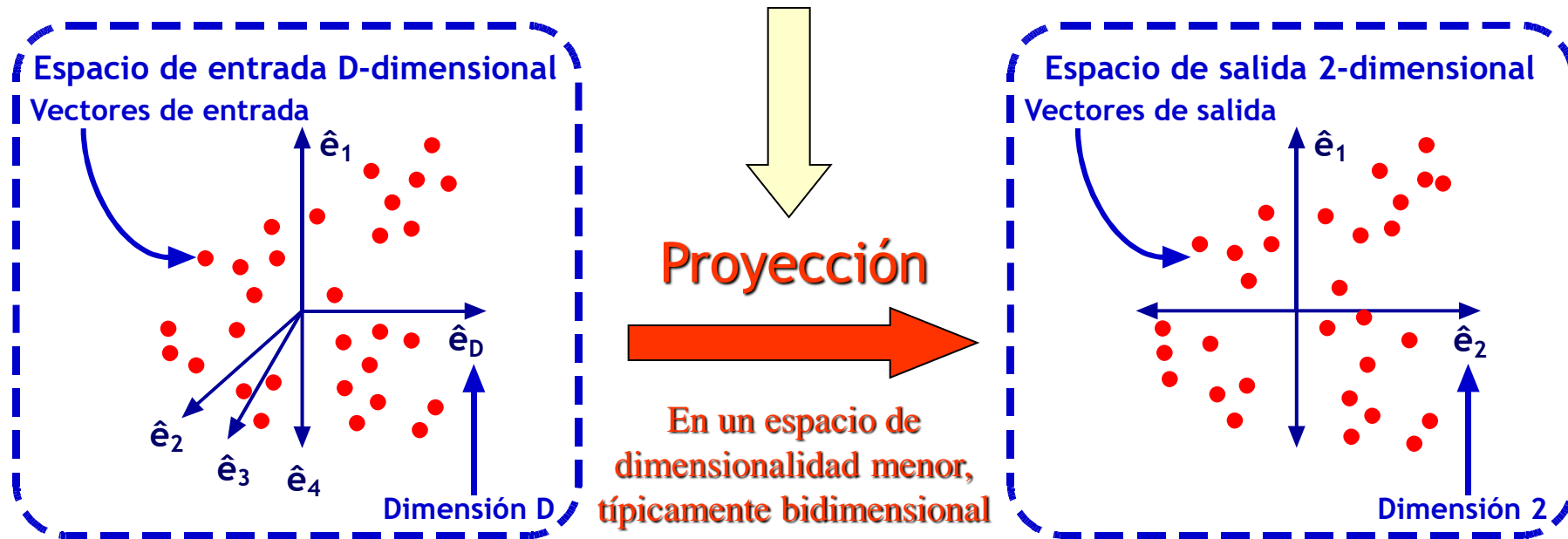
Diagnóstico Médico

Mejorar Experiencia de Clientes

Segmentación de tiendas, E-commerce.

Visualización de Datos Multidimensionales

Dificultad en la visualización directa de vectores con más de 3 atributos

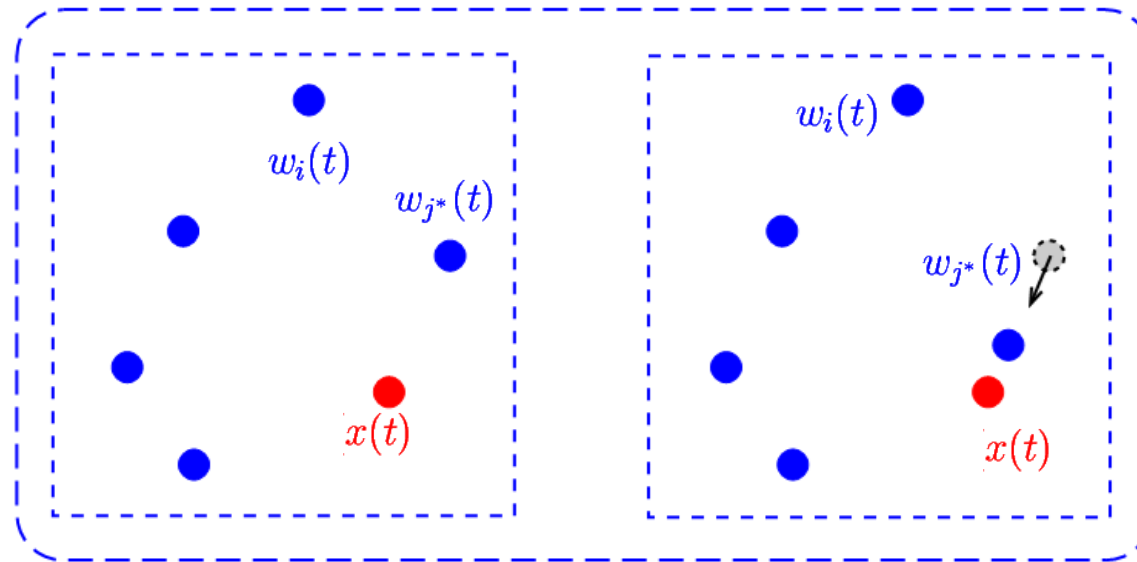


Las proyecciones deben preservar las distancias y la topología entre el espacio de entrada y el espacio de salida, tanto como sea posible.

Principio Funcionamiento Algoritmos Clustering

$$j^* = \operatorname{argmin}_{i=1,\dots,N} \|x(t) - w_i\|,$$

Centroide Ganador

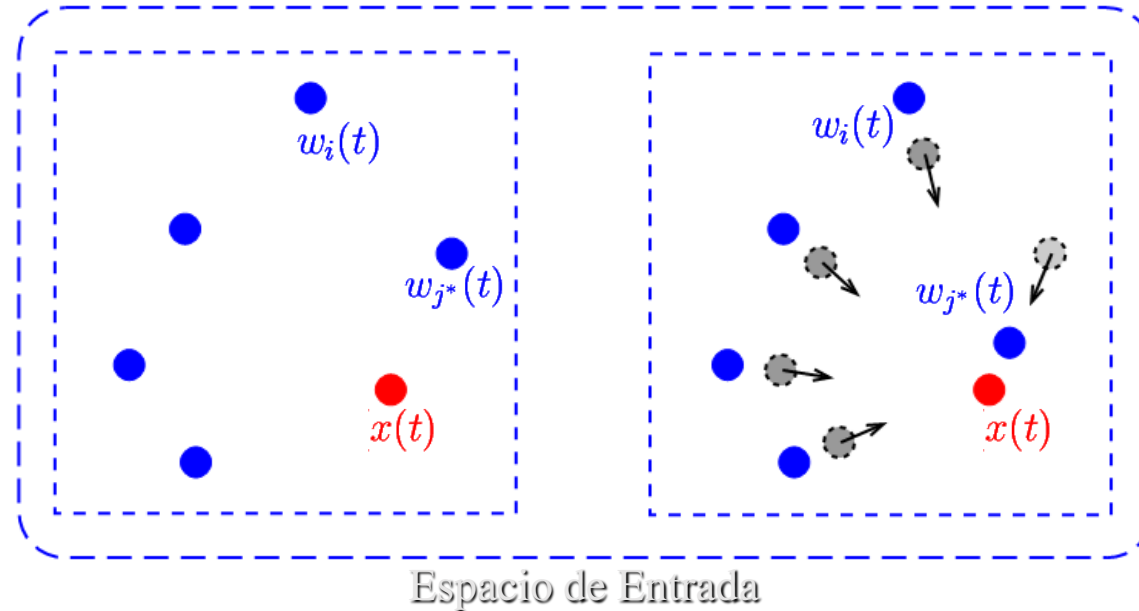


Espacio de Entrada

$$\Delta w_{j^*}(t) = w_{j^*}(t+1) - w_{j^*}(t) = \alpha(t)[x(t) - w_{j^*}(t)],$$

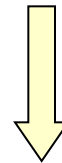
**Ley de Ajuste Auto-Organizativa Fuerte
de los Vectores Centroides**

Principio Funcionamiento Algoritmos Clustering



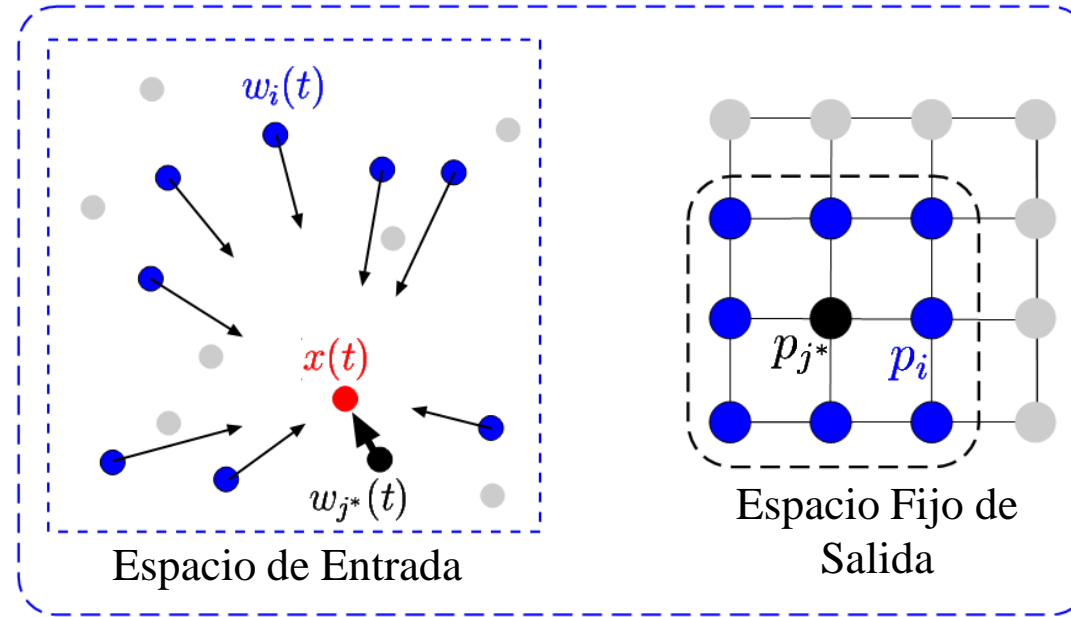
$$\Delta w_i(t) = w_i(t+1) - w_i(t) = \alpha(t)h(t)[x(t) - w_i(t)],$$

**Ley de Ajuste Auto-Organizativa Débil
de los Vectores Centroides**



Función Vecindad

Mapa Auto-Organizativo de Kohonen



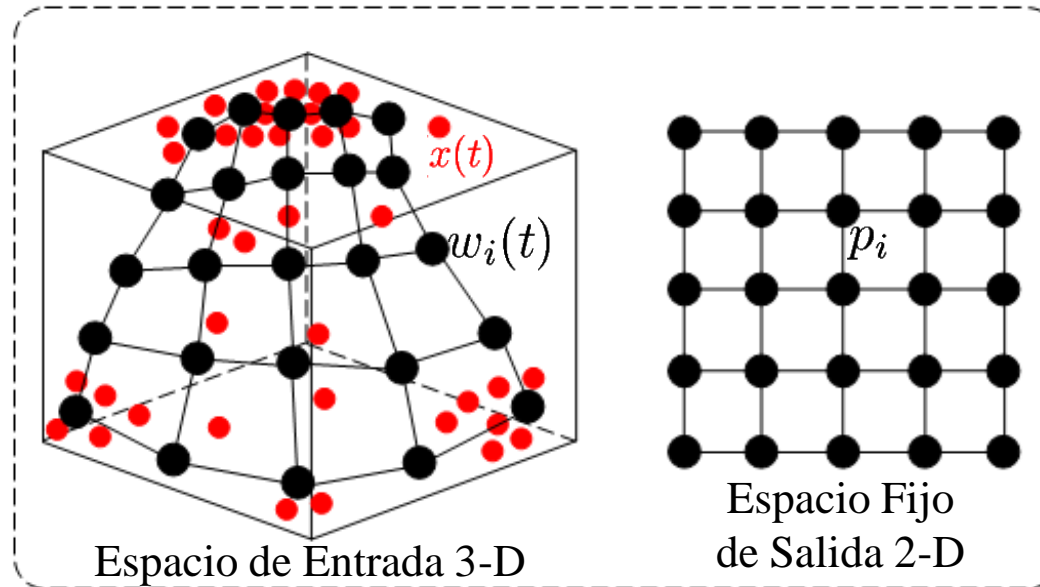
$$\Delta w_i(t) = w_i(t+1) - w_i(t) = \alpha(t) h_{ij^*}(t) [x(t) - w_i(t)],$$

$$h_{ij^*}(t) = e^{-\frac{\|p_i - p_{j^*}\|^2}{L(t)^2}}, \quad L(t) = L_i \left(\frac{L_f}{L_i} \right)^{\left(\frac{t}{t_{max}} \right)}$$

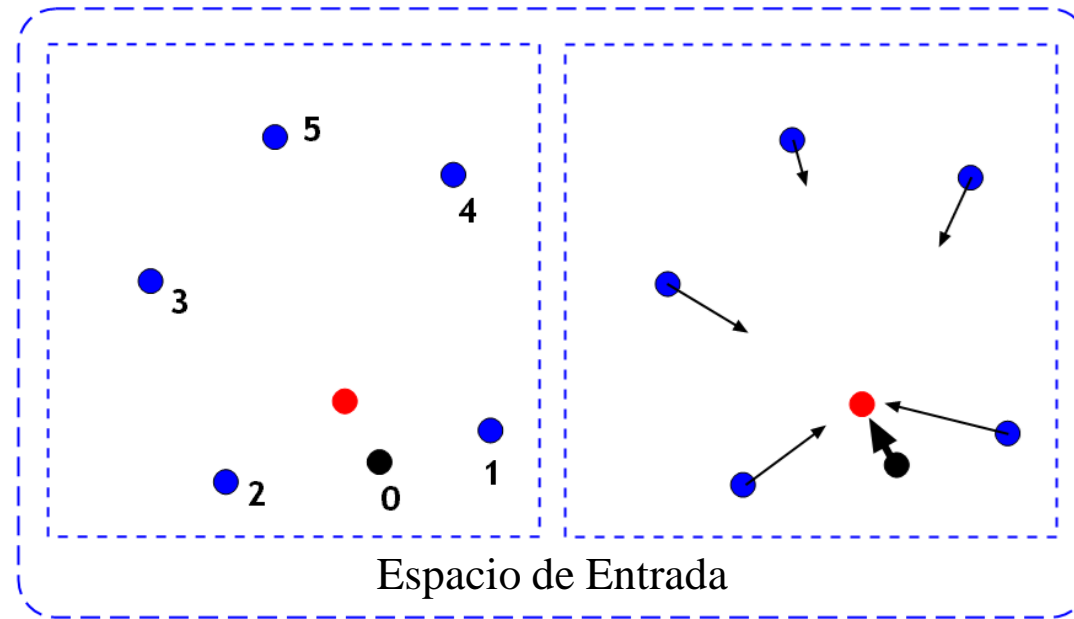
$$\alpha(t) = \alpha_i \left(\frac{\alpha_f}{\alpha_i} \right)^{\left(\frac{t}{t_{max}} \right)}$$

Mapa Auto-Organizativo de Kohonen (2)

- **Algoritmo SOM**
 1. Inicializar los vectores centroides
 2. Presentar un vector de entrada a la red
 3. Encontrar la unidad ganadora
 4. Actualizar los vectores centroides
 5. Aumentar el contador de las iteraciones
 6. Volver al punto 2



Variantes: Gas Neuronal (NG)

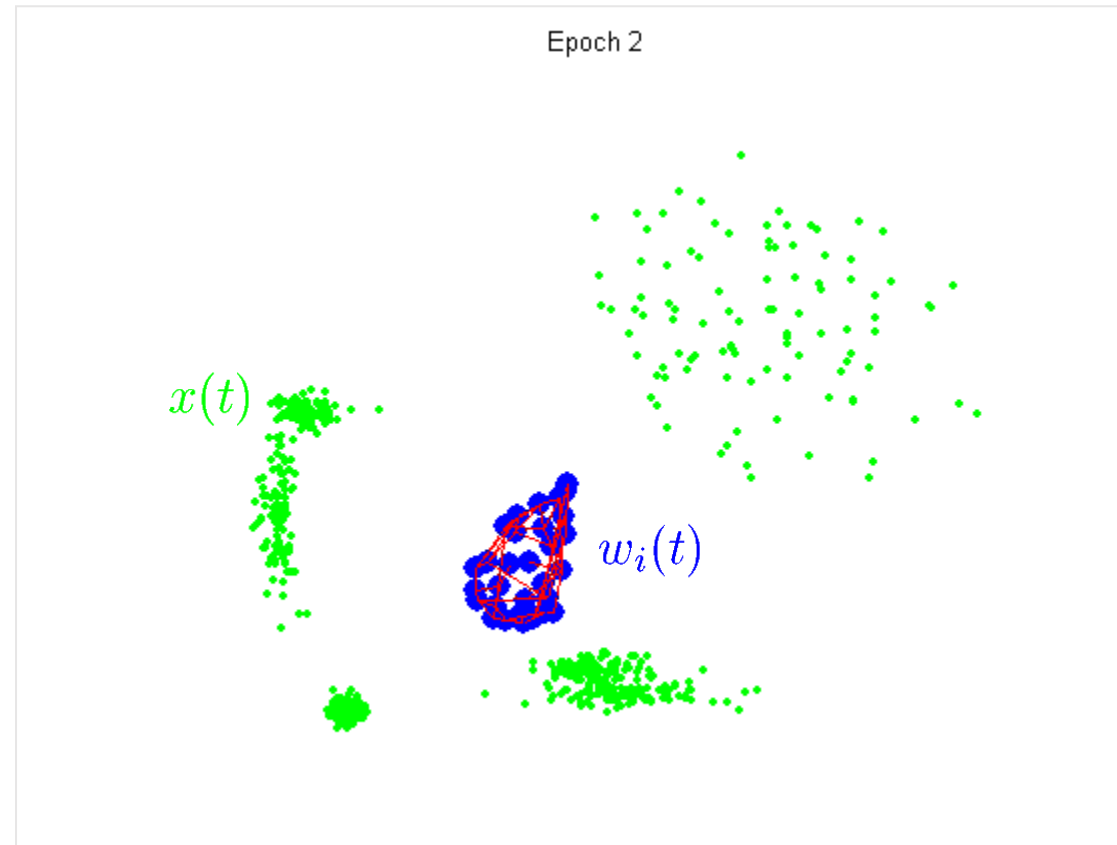


$$\Delta w_i(t) = \epsilon(t) h_\lambda(t) (x(t) - w_i(t)),$$

$$h_\lambda(t) = e^{-\frac{k}{\lambda(t)}} \quad \lambda(t) = \lambda_i \left(\frac{\lambda_f}{\lambda_i} \right)^{\left(\frac{t}{t_{max}} \right)}$$

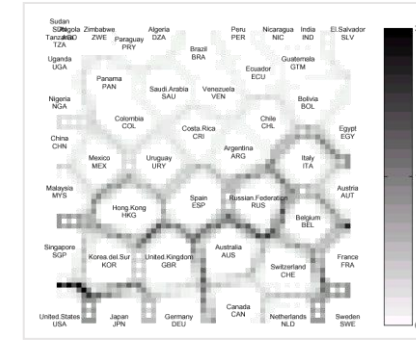
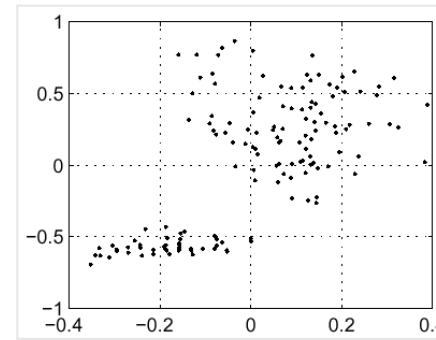
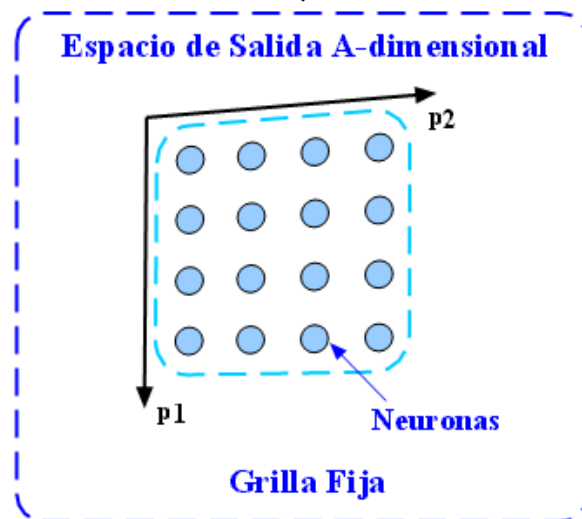
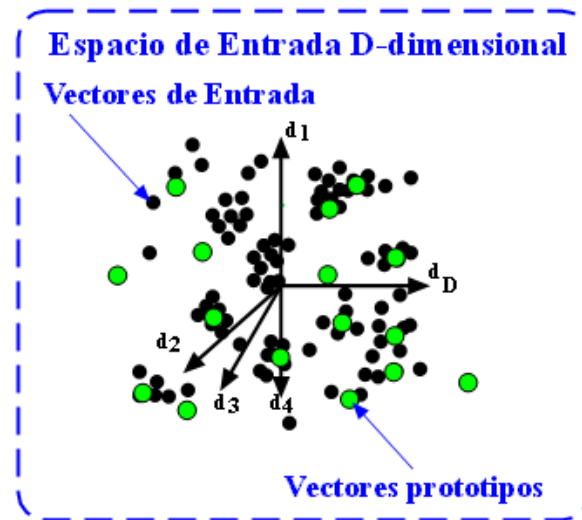
$$\epsilon(t) = \epsilon_i \left(\frac{\epsilon_f}{\epsilon_i} \right)^{\left(\frac{t}{t_{max}} \right)}$$

Variantes: Gas Neuronal (2)



Demostración del algoritmo NG para un conjunto de entrada formado por *clusters* bidimensionales. En el video se aprecia la actualización de los vectores centroides.

Esquemas de Visualización Existentes



Esquemas de Visualización *On-line*

DIPOL-SOM

Double SOM (DSOM)

Coordenadas Adaptivas (AC)

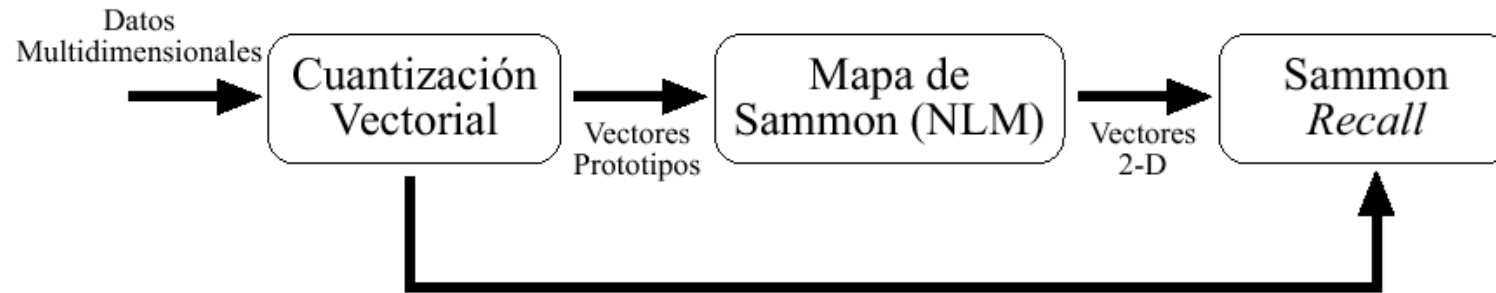
Esquemas de Visualización *Off-line*

Matriz-U

Matriz-P

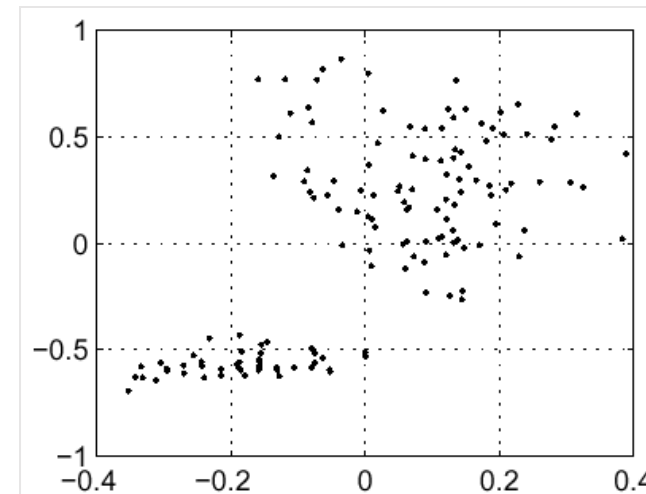
Mapa de Sammon (NLM)

SOM / Mapa de Sammon (NLM)



$$E = \frac{1}{\sum_{j=1}^N \sum_{i=1}^j d_{ij}^w} \sum_{j=1}^N \sum_{i=1}^j \frac{[d_{ij}^w - d_{ij}^p]^2}{d_{ij}^w}$$
$$y_{il}(t+1) = y_{il}(t) - \xi \frac{\frac{\partial E(t)}{\partial y_{il}(t)}}{\left| \frac{\partial^2 E(t)}{\partial y_{il}(t)^2} \right|}$$

- No proporciona una función de mapeo
- Carga Computacional $O(N^2)$
- Estrategia de Optimización Cuadrática
- No utiliza la grilla de salida del SOM



DIPOL – SOM

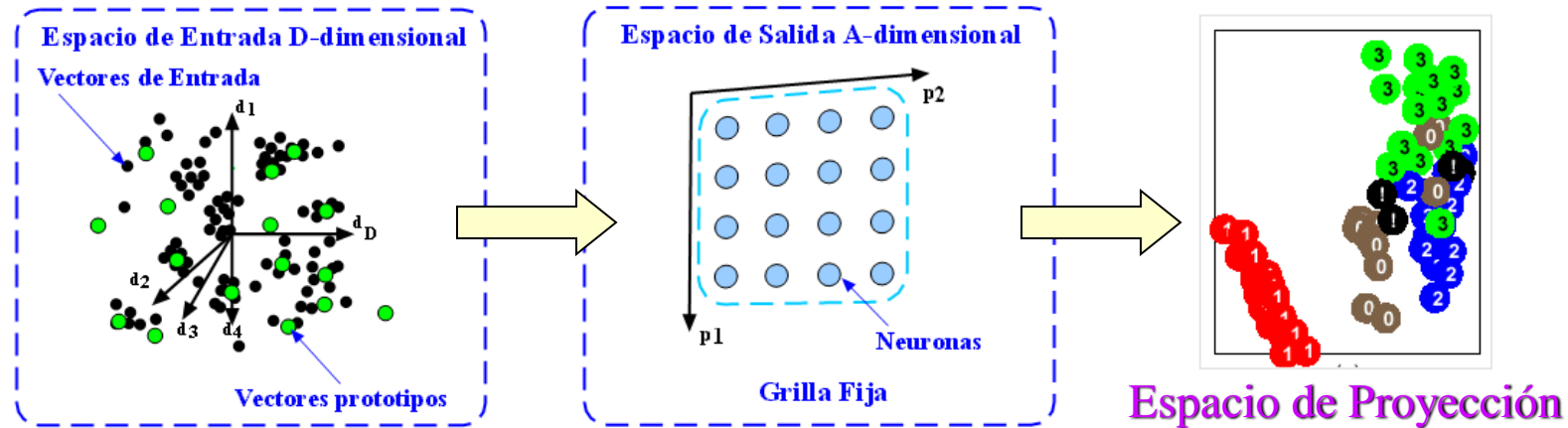
- Posee una capa de proyección adicional a la de SOM
- Vectores posición son actualizados
- Dos modos de entrenamiento: *off-line* y *on-line*
- Función vecindad medida en el espacio de salida de SOM

$$d_{ij}^w = \|w_i - w_{j^*}\| \quad Dir = \text{sgn}(d_{ij}^p - d_{ij}^w) \quad d_{ij}^p = \|p_i - p_{j^*}\|$$

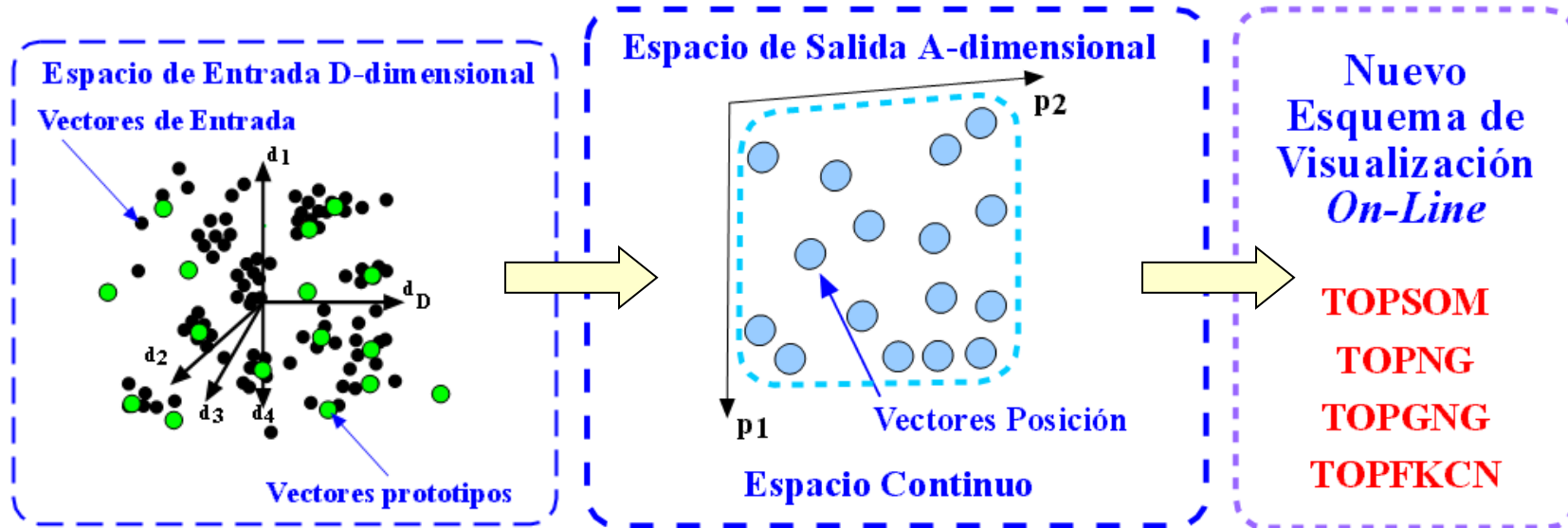


$$\Delta p_i = Dir \times \alpha^p(t) \times H_{ij}^p(t) \times (p_{j^*}(t) - p_i(t))$$

Ley de Ajuste de los Vectores Posición



Proyecciones TOpológicas (TOP)



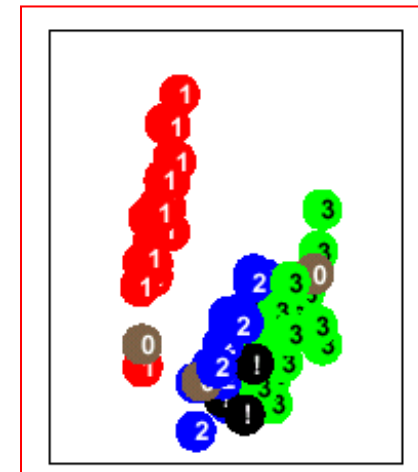
Mapa Auto-Organizativo (SOM)
 Gas Neuronal (NG)
 Gas Neuronal Constructivo (GNG)
 Fuzzy Kohonen C-Means (FKCN)

$$D_{ij*} = \|p_i - p_{j*}\|$$

$$d_{ij*} = \|w_i - w_{j*}\|$$

$$\Delta p_i(t) = \alpha_2(t) (p_j^*(t) - p_i(t)) (\lambda \cdot D_{ij*} - d_{ij*}),$$

Ley de Ajuste Esquema TOP



Medidas de Desempeño

Preservación de Distancias

$$E = \frac{1}{\sum_{j=1}^N \sum_{i=1}^j d_{ij^w}} \sum_{j=1}^N \sum_{i=1}^j \frac{[d_{ij^w} - d_{ij^p}]^2}{d_{ij^w}}$$

Error o *stress* de
Sammon

Preservación de Topología

$$P_3(k, i) = \left(\prod_{j=1}^k \frac{d^{G_V}(w_i, w_{n_j^A(i)})}{d^{G_V}(w_i, w_{n_j^V(i)})} \bullet \frac{d^A(p_i, p_{n_j^A(i)})}{d^A(p_i, p_{n_j^V(i)})} \right)^{\frac{1}{2k}}$$

$$P_m = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{k=1}^{N-1} \log(P_3(k, i))$$

Producto
Topográfico
modificado

Medidas de Desempeño (2)

Preservación de Topología

$$q_{m_{ji}} = \begin{cases} 3, & \text{if } NN_{ji^w} = NN_{ji^p} \\ 2, & \text{if } NN_{ji^w} = NN_{jl^p}, l \in [1, n], i \neq l \\ 1, & \text{if } NN_{ji^w} = NN_{jt^p}, t \in [n, k], n < k \\ 0, & \text{otro caso.} \end{cases}$$

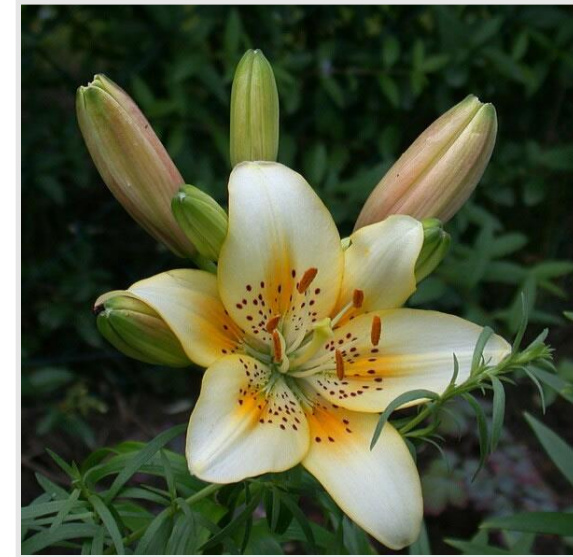
$$q_m = \frac{1}{3n \times N} \sum_{j=1}^N \sum_{i=1}^n q_{m_{ji}}$$

Medida q_m

Bases de Datos

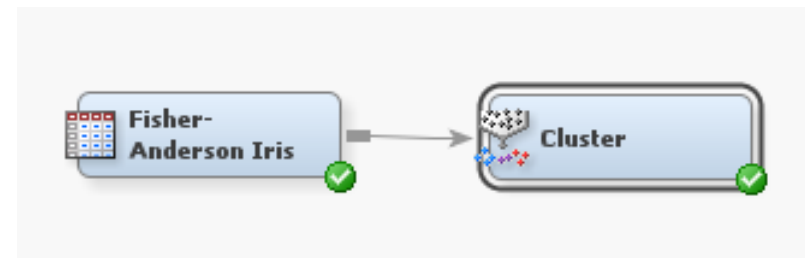
■ Base de Datos Iris

- *Benchmark*
- 150 ejemplos con 4 atributos
- 3 clases con 50 ejemplos cada una:
 - **Iris Setosa**
 - **Iris Virgínica**
 - **Iris Versicolor**



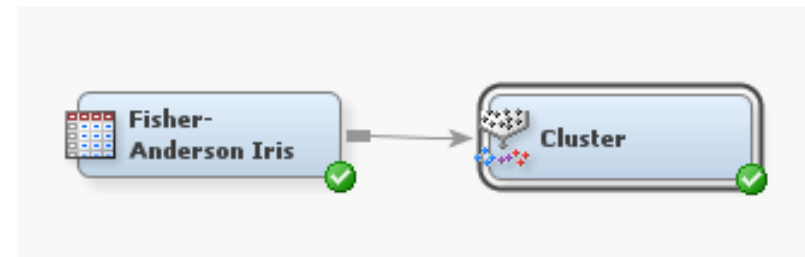
Flor de Lis

| General | |
|---|-----------------------------|
| ID de nodo | Clus |
| Datos importados | ... |
| Datos exportados | ... |
| Notas | ... |
| Entrenamiento | |
| Variables | ... |
| Rol de variable cluster | Segmento |
| Estandarización interna | Estandarización |
| <input type="checkbox"/> Número de clusters | |
| Método de especificación | Especificado por el usuario |
| Número máximo de clusters | 10 |
| <input type="checkbox"/> Criterio de selección | |
| Método cluster | Ward |
| Máximo preliminar | 50 |
| Mínimo | 2 |
| Máximo final | 20 |
| Corte CCC | 3 |
| <input type="checkbox"/> Codificación de las variables de | |
| Codificación ordinal | Rango |
| Codificación nominal | GLM |
| <input type="checkbox"/> Semillas de cluster inicial | |
| Método de inicialización de sem | Predeterminado |
| Radio mínimo | 0.0 |
| Desviación durante entrenamiento | No |
| <input type="checkbox"/> Opciones de entrenamiento | |
| Utilizar predeterminados | Sí |
| Opciones | ... |
| <input type="checkbox"/> Valores ausentes | |
| Variables de intervalo | Predeterminado |
| Variables nominales | Predeterminado |
| Variables ordinales | Predeterminado |
| Método de imputación de scores | Ninguno |
| Puntuación | |
| Rol de variable cluster | Segmento |
| Ocultar variables originales | Sí |
| Editor de etiquetas de cluster | ... |
| Informe | |
| Gráficos del cluster | Sí |
| Perfil del árbol | Sí |
| Gráfico y tabla de distancia | Sí |



| Opciones | |
|--|--------|
| .. Propiedad | Valor |
| Índice de aprendizaje | 0.7 |
| Índice inicial | 0.5 |
| Índice final | 0.02 |
| Número de paso para alcanzar el índice final | 1000 |
| Número máximo de iteraciones | 10 |
| Número máximo de pasos | 1200 |
| Valor del criterio de convergencia | 1.0E-4 |
| Índice de aprendizaje Índice de aprendizaje. | |
| <div>Aceptar</div> <div>Cancelar</div> | |

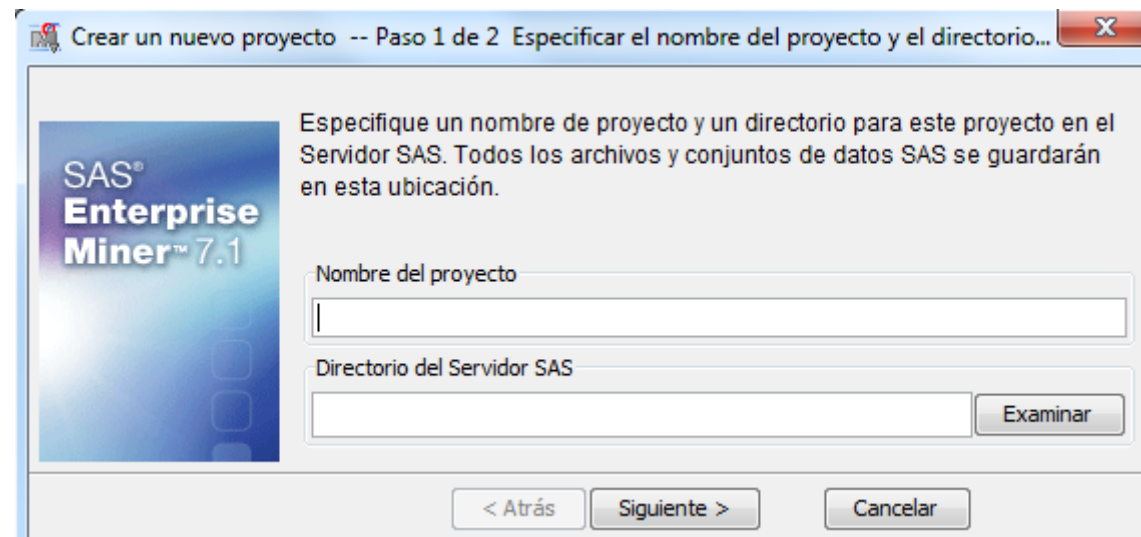
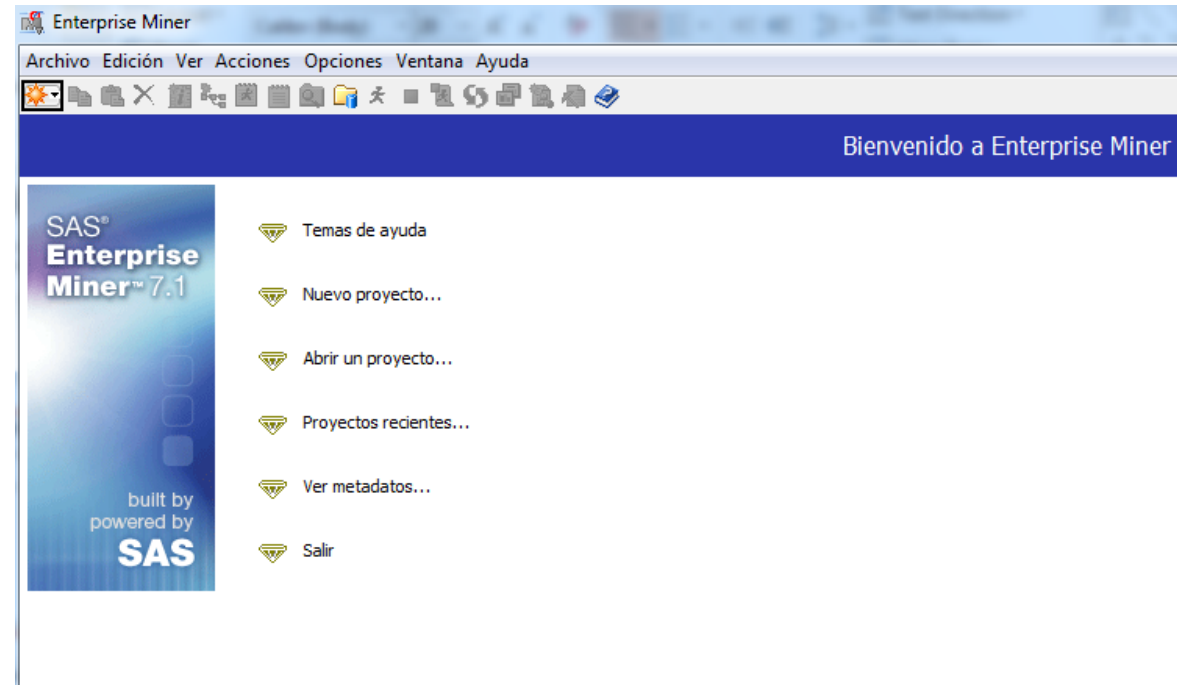
| General | |
|---|-----------------------------|
| ID de nodo | Clus |
| Datos importados | ... |
| Datos exportados | ... |
| Notas | ... |
| Entrenamiento | |
| Variables | ... |
| Rol de variable cluster | Segmento |
| Estandarización interna | Estandarización |
| <input type="checkbox"/> Número de clusters | |
| Método de especificación | Especificado por el usuario |
| Número máximo de clusters | 10 |
| <input type="checkbox"/> Criterio de selección | |
| Método cluster | Ward |
| Máximo preliminar | 50 |
| Mínimo | 2 |
| Máximo final | 20 |
| Corte CCC | 3 |
| <input type="checkbox"/> Codificación de las variables de | |
| Codificación ordinal | Rango |
| Codificación nominal | GLM |
| <input type="checkbox"/> Semillas de cluster inicial | |
| Método de inicialización de sem | Predeterminado |
| Radio mínimo | 0.0 |
| Desviación durante entrenamiento | No |
| <input type="checkbox"/> Opciones de entrenamiento | |
| Utilizar predeterminados | Sí |
| Opciones | ... |
| <input type="checkbox"/> Valores ausentes | |
| Variables de intervalo | Predeterminado |
| Variables nominales | Predeterminado |
| Variables ordinales | Predeterminado |
| Método de imputación de scores | Ninguno |
| Puntuación | |
| Rol de variable cluster | Segmento |
| Ocultar variables originales | Sí |
| Editor de etiquetas de cluster | ... |
| Informe | |
| Gráficos del cluster | Sí |
| Perfil del árbol | Sí |
| Gráfico y tabla de distancia | Sí |



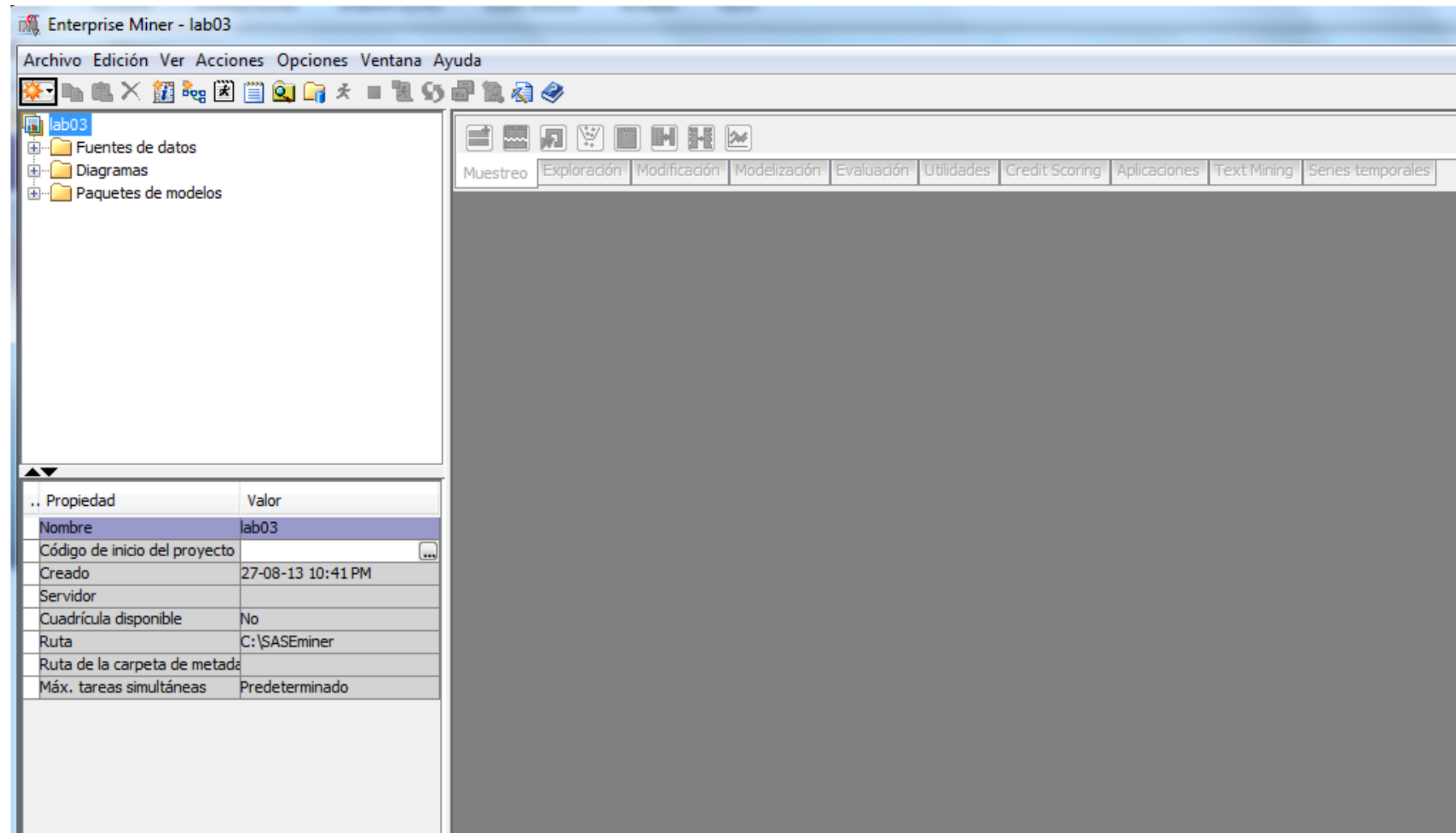
| Opciones | |
|--|--------|
| .. Propiedad | Valor |
| Índice de aprendizaje | 0.7 |
| Índice inicial | 0.5 |
| Índice final | 0.02 |
| Número de paso para alcanzar el índice final | 1000 |
| Número máximo de iteraciones | 10 |
| Número máximo de pasos | 1200 |
| Valor del criterio de convergencia | 1.0E-4 |
| Índice de aprendizaje Índice de aprendizaje. | |
| <div>Aceptar</div> <div>Cancelar</div> | |

Iniciar SAS Enterprise Miner

- Seguir las instrucciones para ejecutar la aplicación SAS Enterprise Miner en el PC.
- Pinchar Nuevo proyecto
- Nombre del proyecto: DIE_Lab_202101
- Directorio del Servidor SAS: Dejar el default
- Siguiente
- Finalizar



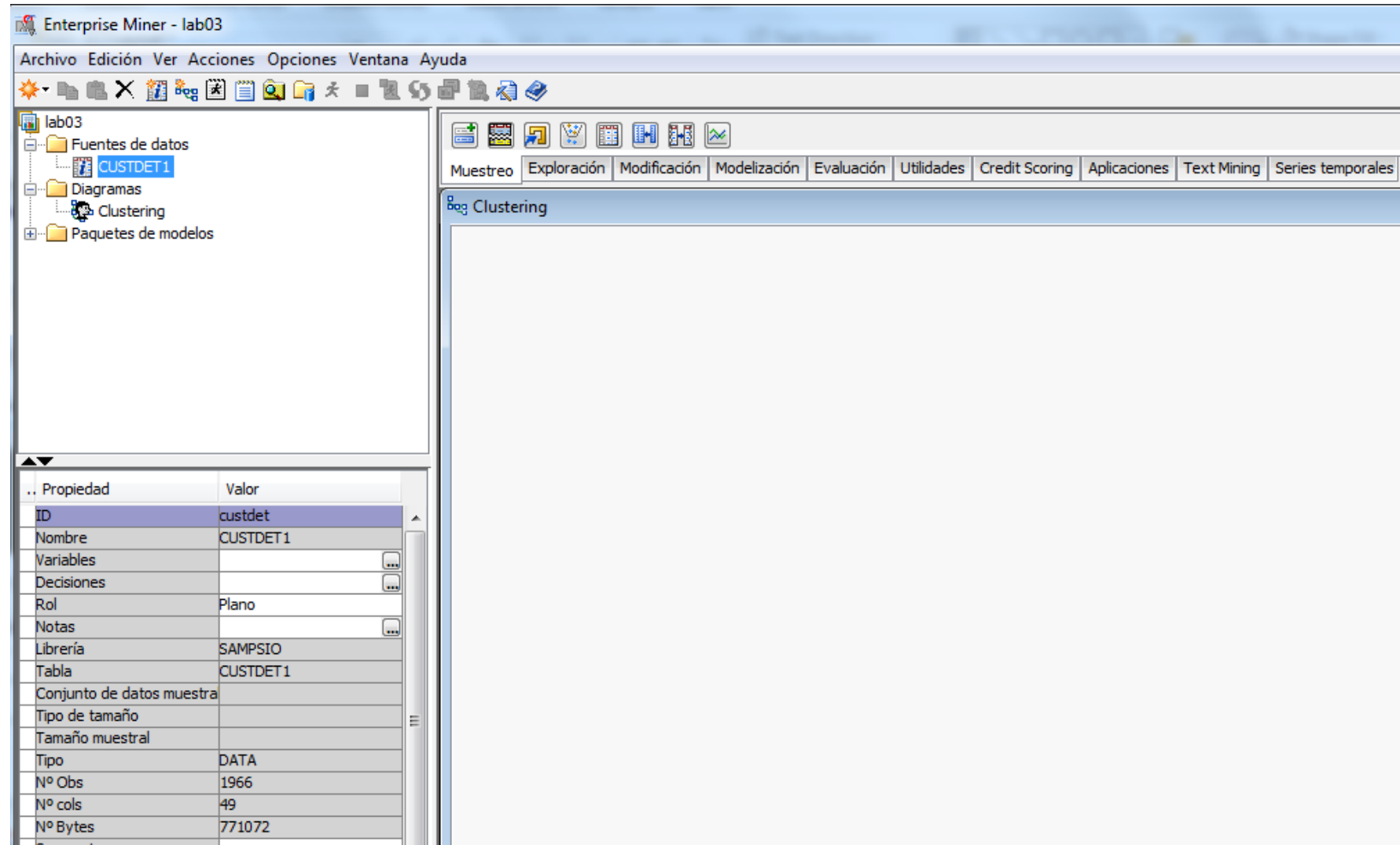
Conociendo la interfaz



Crear Diagrama y Datos

- Sobre Diagramas, botón derecho y crear un diagrama
- Nombrar Clustering al diagrama.
- Aceptar
- Sobre Fuentes de Datos, botón derecho y crear una fuente de datos
- Siguiente
- Examinar
- Seleccionar librería sampsis
- Seleccionar archivo llamado Custdet1 en la columna nombre.
- Presionar Aceptar.
- Presionar Siguiente 6 veces
- Presionar Finalizar

Proyecto Creado



Crear flujo

- Arrastrar con el mouse el nodo fuente de datos Custdet1 a la ventana del diagrama o «área de dibujo»
- De las carpetas ubicadas «arriba del área de dibujo» pinchar sobre Exploración/Explore y arrastrar con el mouse el nodo Cluster a la ventana del diagrama o «área de dibujo»
- Nota: SAS sigue la metodología llamada SEMMA (Sample, Explore, Modify, Model, Assess) para llevar a cabo proyectos de Data Mining (Proceso KDD). En SAS Enterprise Miner existe una carpeta con nodos especializados para cada uno de los pasos de esta metodología.
- Conectar el nodo fuente de datos Custdet1 al nodo Cluster acercando el mouse al nodo Custdet1, presionando y arrastrando hasta el nodo Cluster.

Flujo Creado

The screenshot displays the Enterprise Miner - lab03 application window. The interface includes a menu bar (Archivo, Edición, Ver, Acciones, Opciones, Ventana, Ayuda), a toolbar, and a left-hand tree view showing the project structure: lab03, Fuentes de datos, CUSTDET1, Diagramas, Clustering, and Paquetes de modelos. The main workspace shows a workflow diagram with a 'CUSTDET1' data source node connected to a 'Cluster' model node. The bottom-left panel displays the properties for the selected 'Cluster' node, organized into sections: General, Entrenamiento, and Criterio de selección.

| Propiedad | Valor |
|------------------------------|-----------------|
| General | |
| ID de nodo | Clus |
| Datos importados | |
| Datos exportados | |
| Notas | |
| Entrenamiento | |
| Variables | |
| Rol de variable cluster | Segmento |
| Estandarización interna | Estandarización |
| Número de clusters | |
| Método de especificación | Automático |
| Número máximo de clusters | 10 |
| Criterio de selección | |
| Método cluster | Ward |
| Máximo preliminar | 50 |
| Mínimo | 2 |

Trabajo sobre el «área de dibujo»

- En el «área de dibujo» presionar botón derecho sobre el nodo Custdet1
- Presionar ejecutar
- Presionar Sí
- Presionar Aceptar
- Aparecerá un símbolo color verde en el extremo inferior derecho del nodo indicando que la ejecución fue exitosa.
- Para explorar los datos presionar botón derecho del mouse sobre el nodo Cluster y seleccionar Editar Variables.
- Aparecerá una ventana con el detalle de cada variable existente en la fuente de datos. Seleccionar las variables que se explorarán manteniendo presionada la tecla CTRL del teclado y eligiendo los nombres de las variables respectivas.

Elegir variables para exploración

Variables - Clus

(ninguno) ☐ no Igual a ...

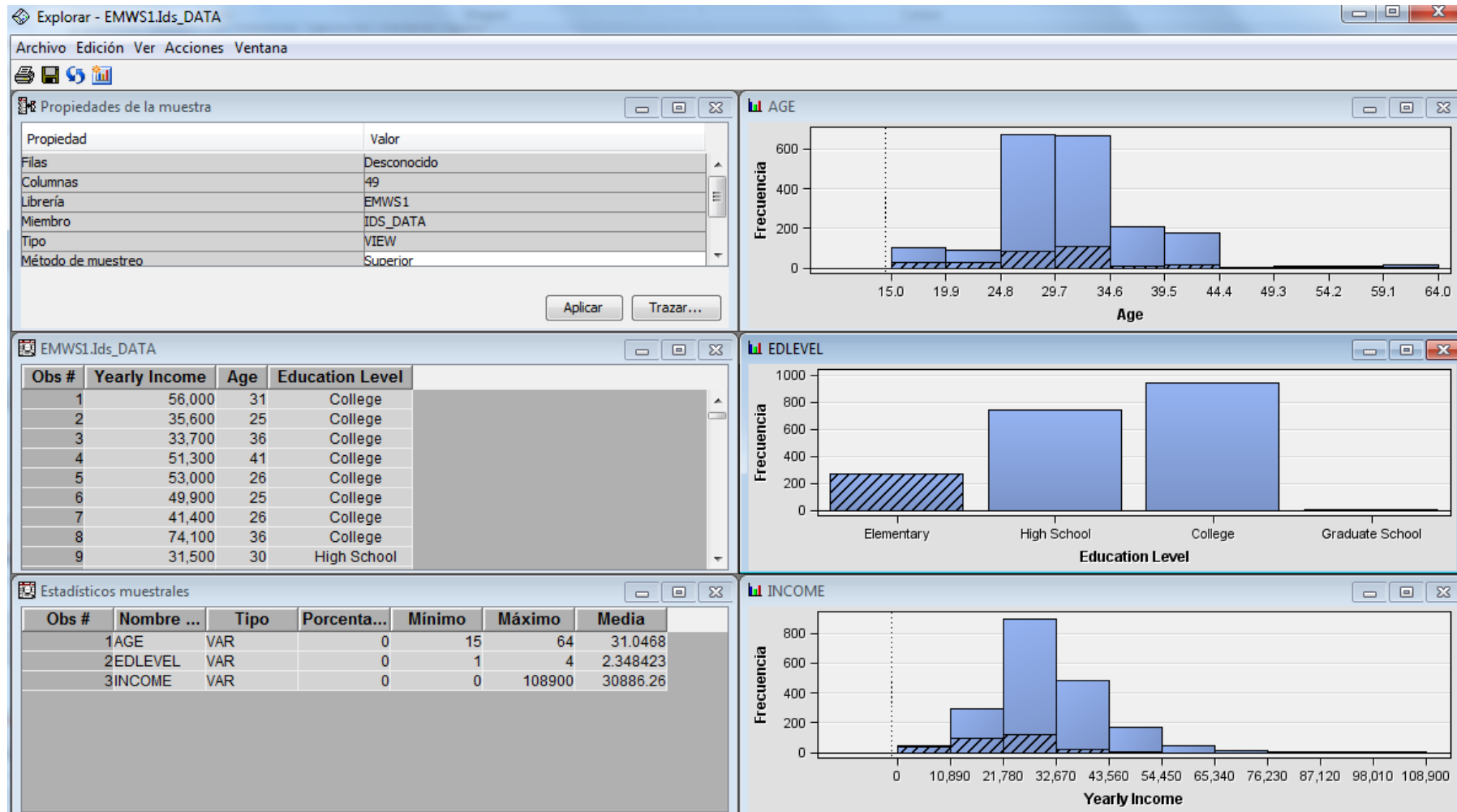
Columnas: ☐ Etiqueta ☐ Mining ☐ Básico

| Nombre | Usar | Informe | Rol | Nivel |
|----------|----------------|---------|---------|-----------|
| ACCTNUM | Predeterminado | No | Entrada | Nominal |
| AGE | Predeterminado | No | Entrada | Intervalo |
| AMOUNT | Predeterminado | No | Entrada | Intervalo |
| APPAREL | Predeterminado | No | Entrada | Intervalo |
| APRTMNT | Predeterminado | No | Entrada | Nominal |
| BLANKETS | Predeterminado | No | Entrada | Intervalo |
| COATS | Predeterminado | No | Entrada | Intervalo |
| COUNTY | Predeterminado | No | Entrada | Intervalo |
| DINING | Predeterminado | No | Entrada | Intervalo |
| DISHES | Predeterminado | No | Entrada | Intervalo |
| DOMESTIC | Predeterminado | No | Entrada | Intervalo |
| EDLEVEL | Predeterminado | No | Entrada | Intervalo |
| FLATWARE | Predeterminado | No | Entrada | Intervalo |
| FREQUENT | Predeterminado | No | Entrada | Intervalo |
| HEAT | Predeterminado | No | Entrada | Intervalo |
| HHAPPAR | Predeterminado | No | Entrada | Intervalo |
| HOMEACC | Predeterminado | No | Entrada | Intervalo |
| HOMEVAL | Predeterminado | No | Entrada | Intervalo |
| INCOME | Predeterminado | No | Entrada | Intervalo |
| JEWELRY | Predeterminado | No | Entrada | Intervalo |
| JOB | Predeterminado | No | Entrada | Intervalo |
| KITCHEN | Predeterminado | No | Entrada | Intervalo |
| LAMPS | Predeterminado | No | Entrada | Intervalo |
| LEISURE | Predeterminado | No | Entrada | Intervalo |
| LINENS | Predeterminado | No | Entrada | Intervalo |
| LUXURY | Predeterminado | No | Entrada | Intervalo |
| MARITAL | Predeterminado | No | Entrada | Intervalo |
| MENSWARE | Predeterminado | No | Entrada | Intervalo |
| MOBILE | Predeterminado | No | Entrada | Intervalo |
| NTITLE | Predeterminado | No | Entrada | Nominal |

Exploración de variables

- Como ejemplo, elegir las variables Age (edad), Edlevel (Nivel educativo) e Income (salario)
- Presionar botón Explorar
- Aparecerá una ventana con histogramas producidos para cada una de las variables seleccionadas
- Todos los histogramas están enlazados. Cuando se elige un intervalo en uno de estos histogramas, los intervalos asociados en los otros histogramas son destacados.
- Ver lo que sucede cuando se elige el intervalo Elementary para la variable Nivel Educativo
- Ver lo que sucede cuando se elige el intervalo College para la variable Nivel Educativo

Histogramas enlazados



Exploración de variables

- Para cambiar el número de intervalos en un histograma dado presionar botón derecho del mouse sobre el histograma (considere como ejemplo el histograma salario)
- Presionar propiedades del histograma (primera opción)
- Seleccionar el número de intervalos (por ejemplo 20 en vez del 10 que viene por default)
- Presionar Aceptar
- Para integrar datos en una ventana de intervalos arrastrar el mouse a través de una ventana de intervalos definida (considere como ejemplo el histograma salario arrastrando el mouse desde 32.670 hasta 43.560)
- Cerrar la ventana de exploración

Elegir variables para cluster

- Volviendo a la ventana donde se listan las variables, elegir todas las variables (desde Acctnum hasta Wcoat) manteniendo presionada la tecla SHIFT del teclado.
- Bajo la columna llamada Usar elegir una de las casillas con nombre Predeterminado y cambiar a No
- Seleccionar bajo la columna Nombre la variable Income (salario)
- Para la variable Income, bajo la columna llamada Usar cambiar a Sí
- Para la variable Numcars (número de automóviles), bajo la columna llamada Usar cambiar a Sí
- Presionar Aceptar
- Presionar botón derecho del mouse sobre el nodo Cluster y ejecutar
- Presionar Sí
- Presionar Resultados...
- La ventana resultados abre con 4 componentes

Entendiendo los resultados

- En la ventana Trazado de Segmento presionar sobre el fragmento color rojo en Segment Variable 5 para la Variable = Income. Se puede ver el siguiente texto:
Segment Variable = 5
Percent (Sum) = 99,56522
Formatted Value = 13613:27225
- Esto significa que en el cluster número 5, el 99,56522% de personas tienen un salario cuyo valor está entre 13613 y 27225 (unidades monetarias).
- Haga un ejercicio similar con los otros segmentos.

Entendiendo los resultados (II)

- Para ver el número de personas en el cluster número 5 presionar Segment N°5 en la ventana Tamaño del Segmento. Se puede ver el siguiente texto:
Segment Id = 5
Frequency of Cluster = 230
- Esto significa que el cluster número 5 incluye 230 personas.
- Explore y entienda los datos

Entendiendo los resultados (III)

- Seleccionar en la parte inferior de la ventana Segment Plot alguno de los colores con Formatted value dado (Como ejemplo, considere Formatted Value = 27225:40838)
- Ahora se puede ver la distribución de Formatted Value para Income (salario) entre 27225 y 40838 unidades monetarias entre diferentes clusters.
- Para ver la Distancia entre Clusters seleccionar Ver → Distancia Cluster → Trazado
- Dé alguna interpretación a la ubicación de los clusters en el plano de visualización.
- Cerrar la ventana Plot
- Cerrar resultados del Cluster

Cambiar parámetros del algoritmo

- En la tabla Propiedad/Valor elegir Número de clusters ➔ Método de especificación. Cambiar Automático por Especificado por usuario.
- En máximo número de clusters colocar el número de clusters que se desea especificar (considere como ejemplo 3)
- Presionar botón derecho del mouse sobre el nodo cluster
- Presionar ejecutar.
- Seleccionar Sí.
- Seleccionar Resultados...

| Propiedad | Valor |
|--|-----------------|
| General | |
| ID de nodo | Clus |
| Datos importados | ... |
| Datos exportados | ... |
| Notas | ... |
| Entrenamiento | |
| Variables | ... |
| Rol de variable cluster | Segmento |
| Estandarización interna | Estandarización |
| Número de clusters | |
| Método de especificación | Automático |
| Número máximo de clusters | 5 |
| Criterio de selección | |
| Método cluster | Ward |
| Máximo preliminar | 50 |
| Mínimo | 2 |
| Máximo final | 20 |
| Corte CCC | 3 |
| Codificación de las variables | |
| Codificación ordinal | Rango |
| Codificación nominal | GLM |
| Semillas de cluster inicial | |
| Método de inicialización de semillas | Predeterminado |
| Radio mínimo | 0.0 |
| Desviación durante entrenamiento | No |
| Opciones de entrenamiento | |
| Utilizar predeterminados | Sí |
| Opciones | ... |
| Valores ausentes | |
| Variables de intervalo | Predeterminado |
| Variables nominales | Predeterminado |
| Variables ordinales | Predeterminado |
| Método de imputación de valores ausentes | Ninguno |
| Puntuación | |
| Rol de variable cluster | Segmento |
| Ocultar variables originales | Sí |
| Editor de etiquetas de cluster | ... |
| Informe | |
| Gráficos del cluster | Sí |
| Perfil del árbol | Sí |
| Trazado y tabla de distancias | Sí |

Cambiar variables de análisis

- En la tabla Propiedad/Valor elegir método Automático nuevamente en Número de Clusters.
- Presionar botón derecho del mouse sobre el nodo cluster
- Seleccionar Editar Variables
- Para la variable Age (edad) bajo la columna Usar cambiar el uso a Sí.
- Seleccionar OK.
- Presionar botón derecho del mouse sobre el nodo cluster
- Presionar ejecutar.
- Seleccionar Sí
- Seleccionar Resultados...
- Analice y comente resultados.

Actividades

Actividad 1

- Encuentre los segmentos de cliente que serían más propensos a aceptar una oferta de autos.
- Proponga 1 set de variables (atributos) para el análisis.
- Analice los resultados para diferente número de clusters.
- Formule sus recomendaciones para gatillar campañas de marketing más precisas, y explique las razones de sus recomendaciones.

Actividades (II)

Actividad 2

- Encuentre los segmentos de clientes que serían más propensos a aceptar una oferta de artículos (defina un conjunto de artículos que los clientes compren) según la cantidad de hijos.
- Proponga 1 set de variables (atributos) para el análisis.
- Analice los resultados para diferente número de clusters.
- Formule sus recomendaciones para gatillar campañas de marketing más precisas, y explique las razones de sus recomendaciones.

Informe

En el informe de resultados incluya:

- 1. Los sets de variables > 3 (atributos) que fueron elegidos para el análisis.
- 2. Número de clusters que fue usado para el análisis para cada set de variables.
- 3. Sus recomendaciones (customer segments) para campañas de marketing según cada combinación de variables.
- 4. Resumen de sus recomendaciones (customer segments) para las campañas de marketing.
- 5. Explicación para las razones de sus recomendaciones.

Listado de variables

| | |
|--|--|
| ACCTNUM – Account Number | LINENS – Linens Purch. |
| AGE - Age | LUXURY – Luxury Items |
| AMOUNT – Dollars Spent | MARITAL – Married (y/n) |
| APPAREL – Apparel Purch. | MENSWARE – Mens Apparel |
| APRTMNT – Rents Apartment | MOBILE – Occupied < 1Year |
| BLANKETS – Blankets Purch. | NTITLE – Name Prefix |
| COATS – Coats Purch. | NUMCARS – Number of Cars |
| COUNTRY – Country Code | NUMKIDS – Number of Kids |
| CUSTDATE – Date 1 st Order | OUTDOOR – Outdoor Prod. |
| DINING – Total Dining | PROMO13 – Promo: 8-13 Months |
| DISHES – Dishes Purch. | PROMO7 – Promo: 1-7 Months |
| DOMESTIC – Domestic Prod. | PURCHASE – Purchase (y/n) |
| EDLEVEL – Education Level | RACE - Race |
| FLATWARE – Flatware Purch. | RECENCY - Recency |
| FREQUET – Order Frequency | RETURN – Total Returns |
| HEAT – Heating Type | SEX - Sex |
| HHAPPAR – His/Her Apparel | STATECOD – State Code |
| HOMEACC – Home Furniture | TELIND – Telemarket Ind. |
| HOMEVAL – Home Value | TMKTORD – Telemarket Ord. |
| INCOME – Yearly Income | TOWELS – Towels Purch. |
| JEWELRY – Jewelry Purch. | TRAVTIME – Travel Time. |
| JOB – Job Category | VALRATIO - \$ Value per Mailing |
| KITCHEN – Kitchen Prod. | WAPPAR – Ladies Apparel |
| LAMPS – Lamps Purch. | WCOAT – Ladies Coats |