

# Técnicas Analíticas de Segmentación de Mercados

Dr. Viterbo H. Berberena González  
Pearson

# El Análisis de Conglomerados con Técnicas Estadísticas

# Objetivos

- Discutir el concepto del método de los centroides.
- Definir medidas de distancia en el análisis de conglomerados.
- Entender los peligros de los grupos forzados.
- Generar un análisis de conglomerados y interpretar los resultados.

# Clasificación no Supervisada

Training Data

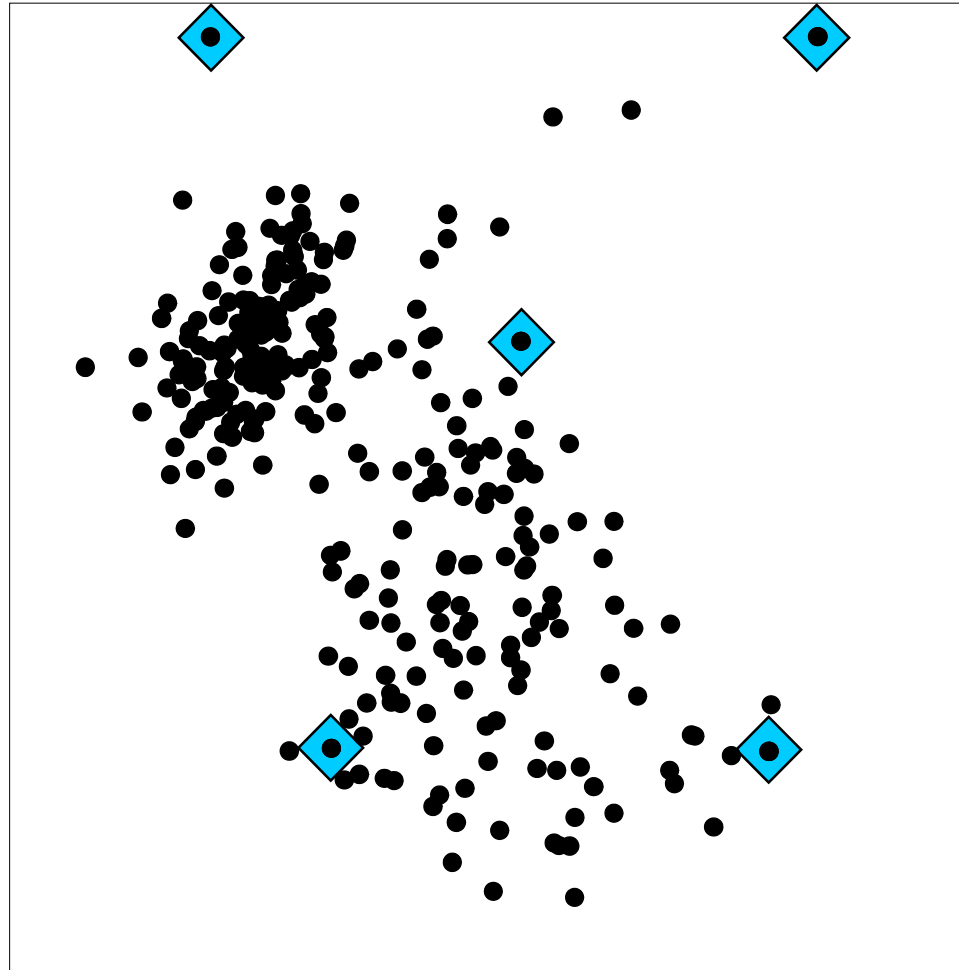
case 1: inputs, ?  
case 2: inputs, ?  
case 3: inputs, ?  
case 4: inputs, ?  
case 5: inputs, ?

Training Data

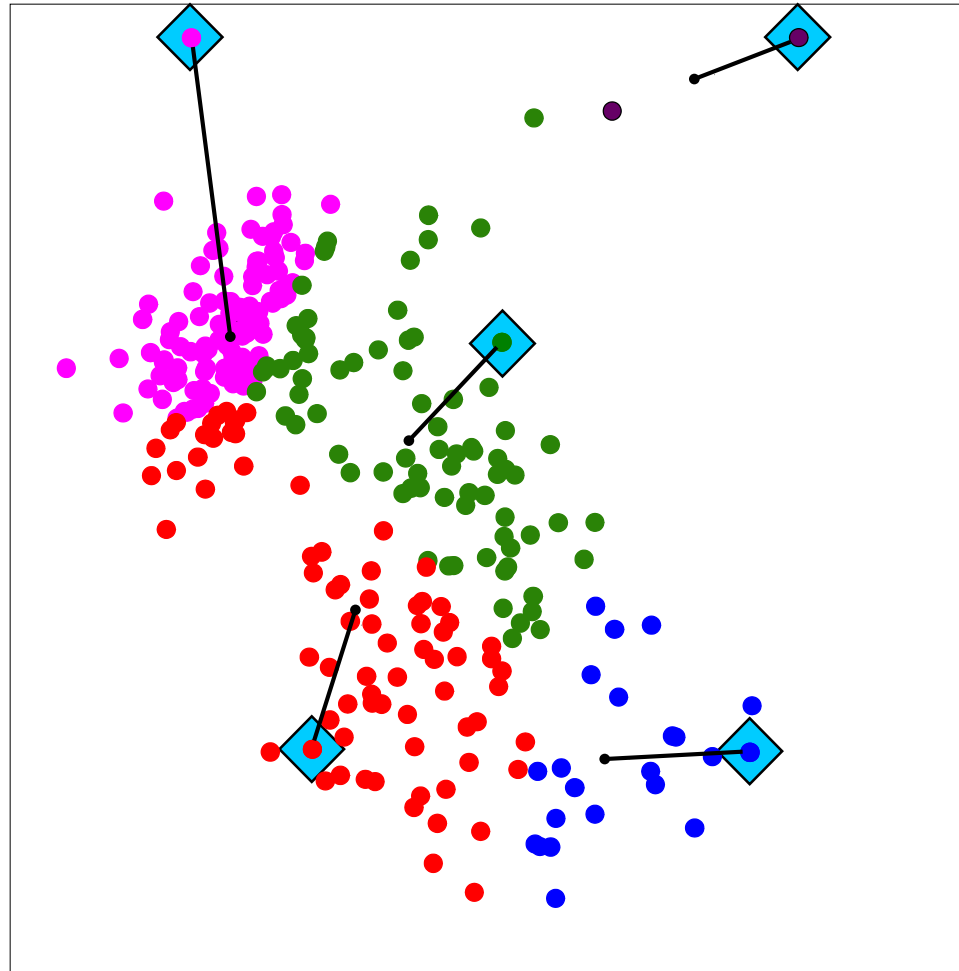
case 1: inputs, cluster 1  
case 2: inputs, cluster 3  
case 3: inputs, cluster 2  
case 4: inputs, cluster 1  
case 5: inputs, cluster 2



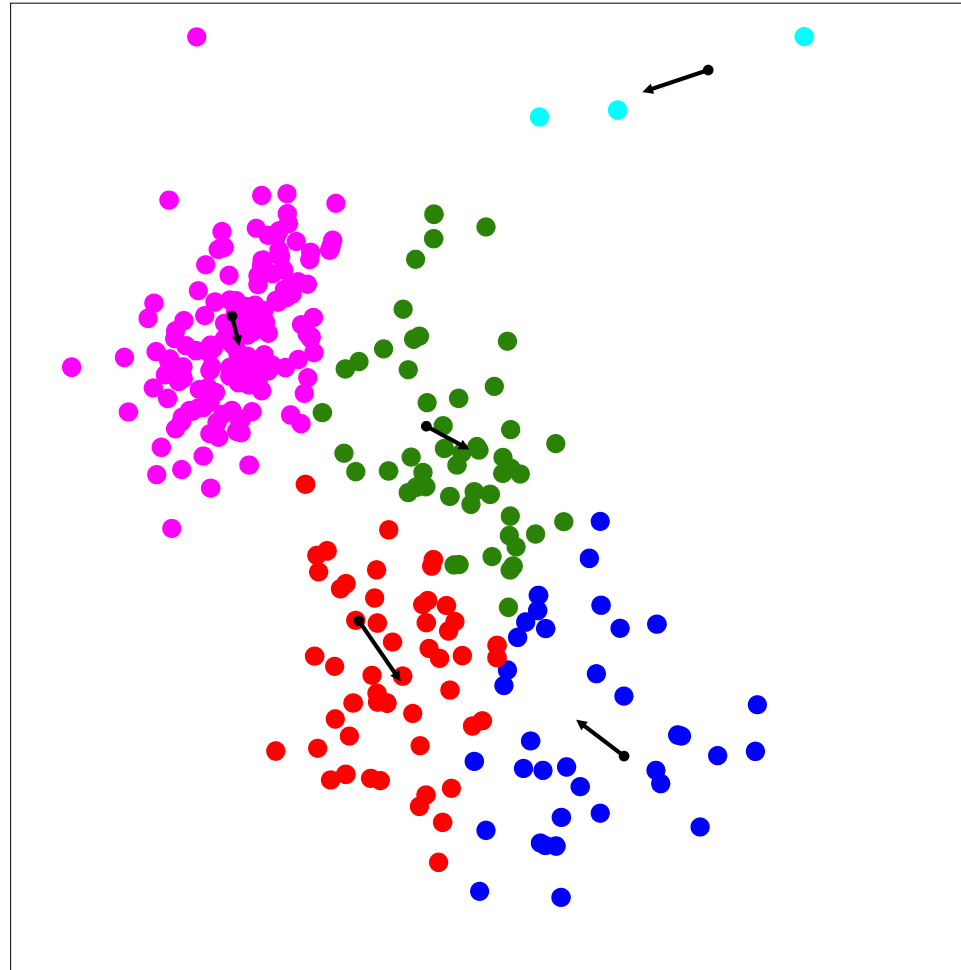
# K-means Clustering



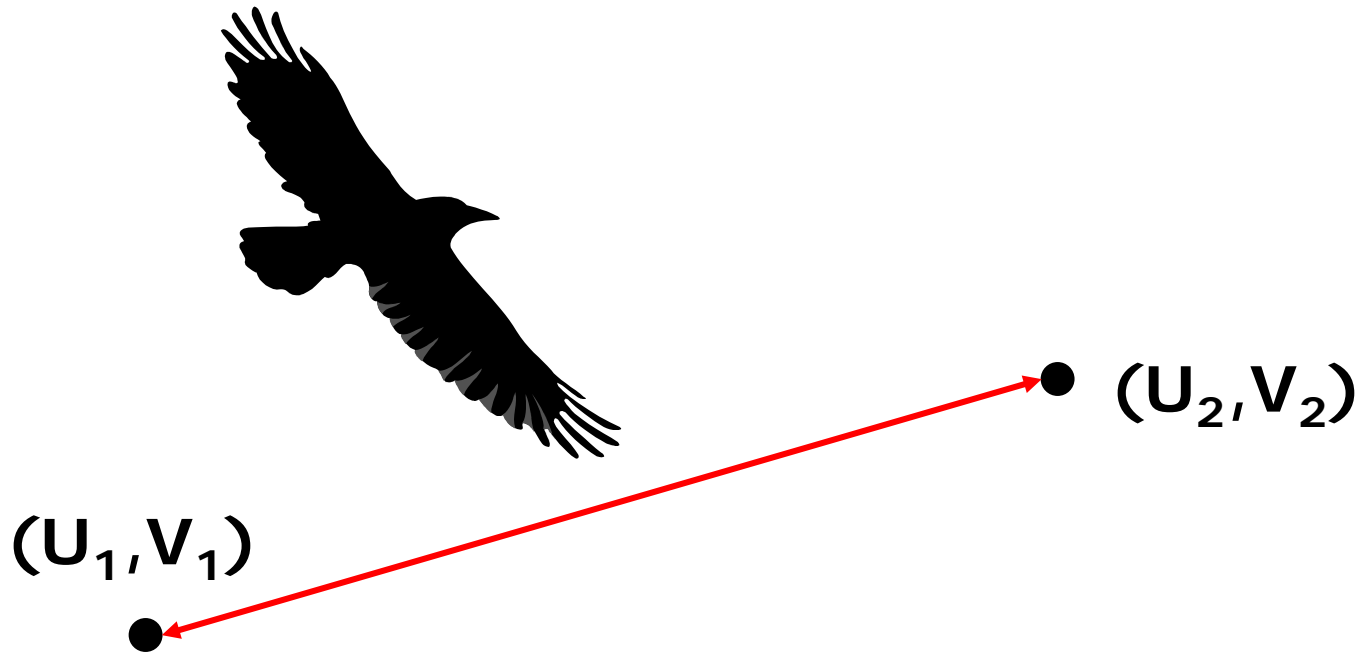
# Asignación de Elementos



# La Reasignación de Elementos

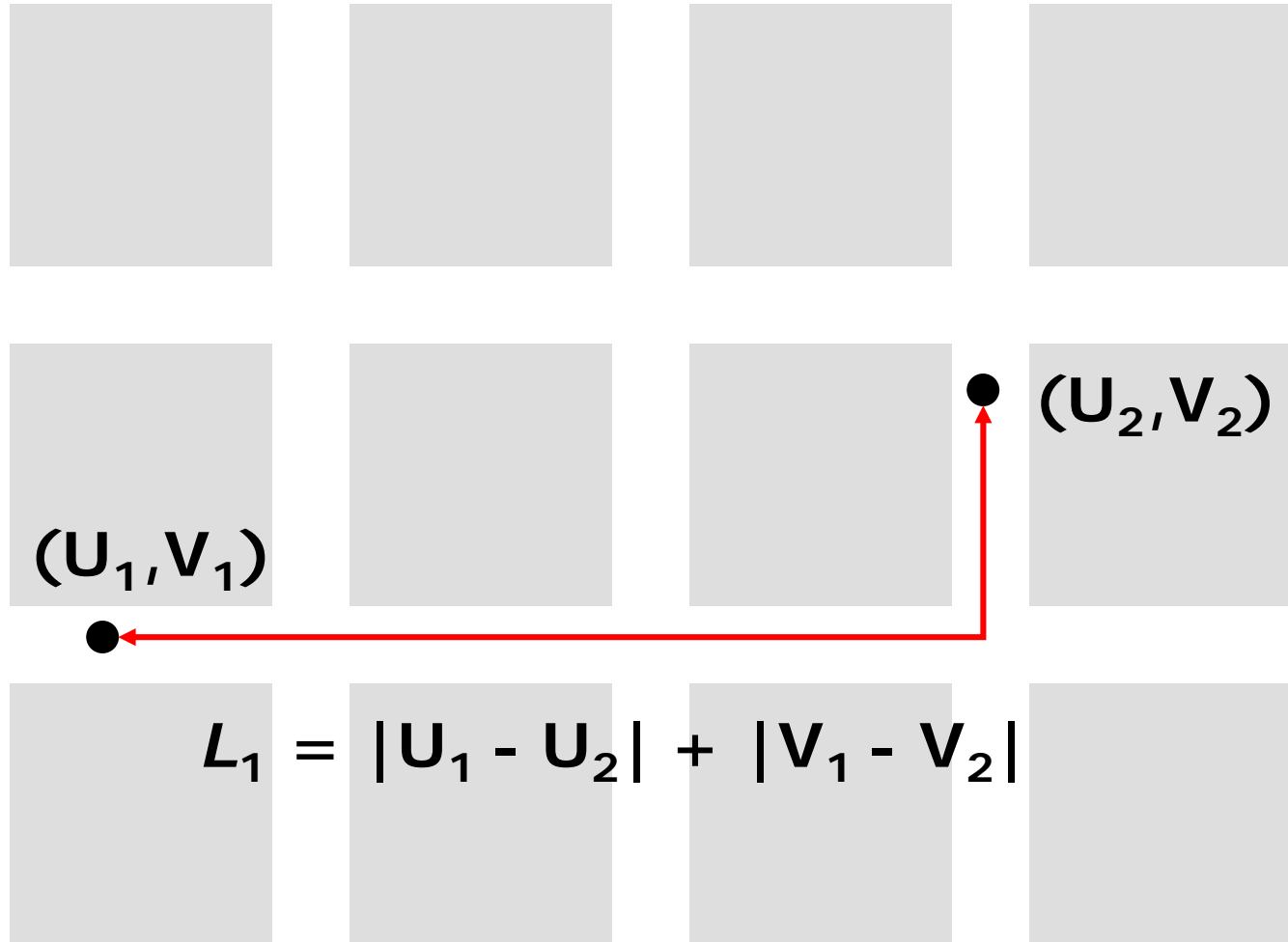


# Distancia Euclidiana

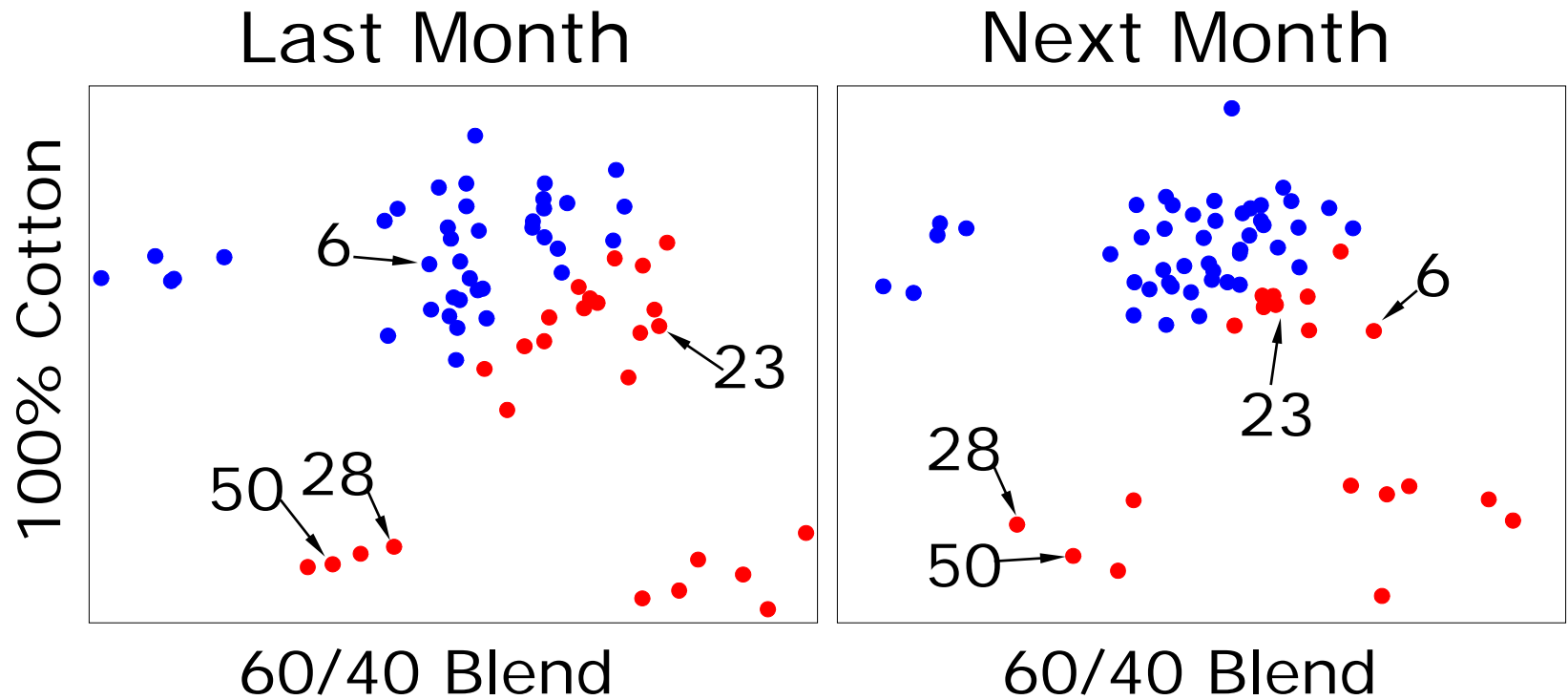


$$L_2 = ((U_1 - U_2)^2 + (V_1 - V_2)^2)^{1/2}$$

# Distancia Manhattan



# Grupos Forzados



# K-Means Clustering Algorithm

Es un algoritmo para particionamiento de  $N$  puntos de una base de datos en  $K$  subconjuntos disjuntos  $S_j$  que minimizan la suma de cuadrados de los puntos  $N_j$ .

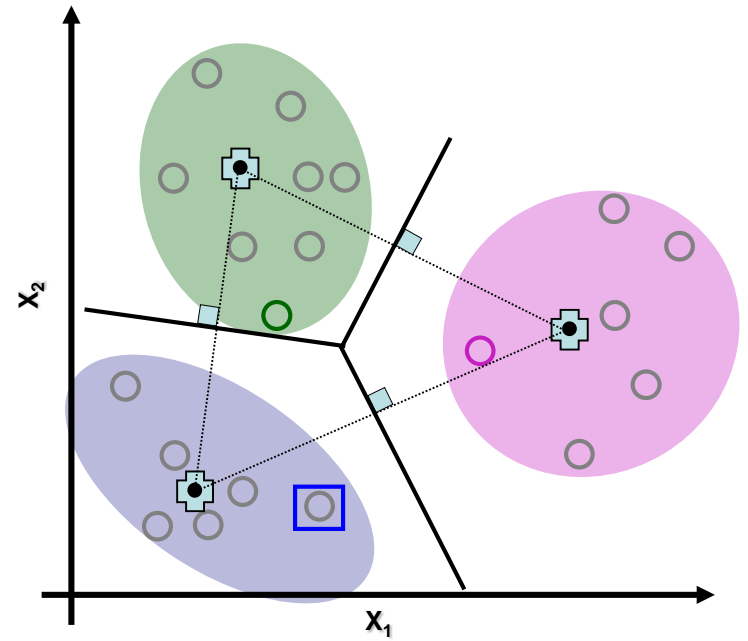
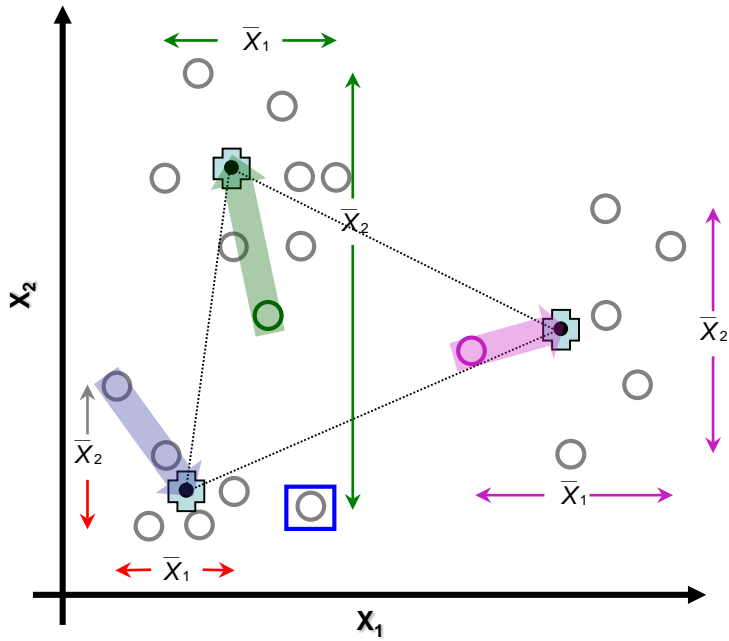
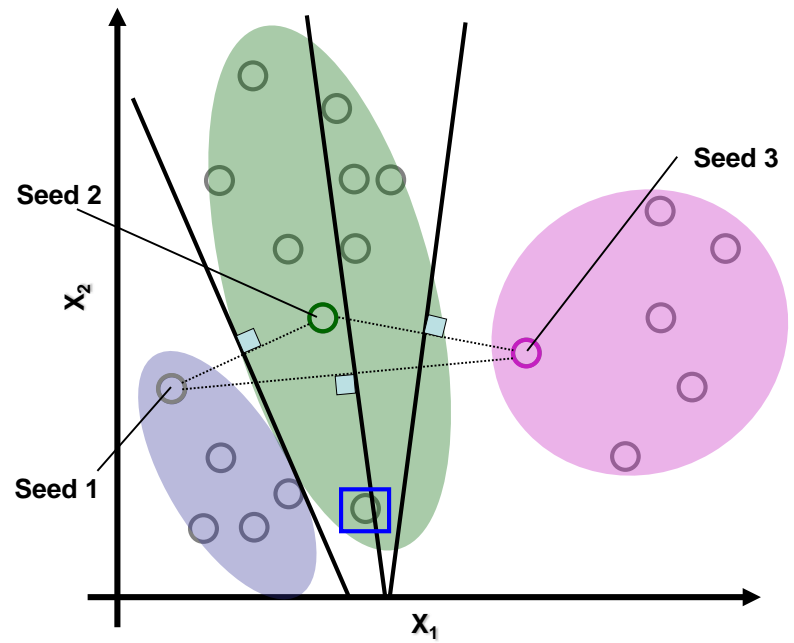
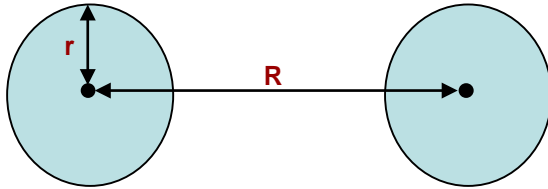
$$J = \sum_{j=1}^K \sum_{n \in S_j} |x_n - \mu_j|^2$$

Donde,

$x_n$  - Es un vector que representa el  $n$ -ésimo punto de datos.

$\mu_j$  - Es el centroide geométrico de los puntos en  $S_j$ .

# Detección Automática de Clusters

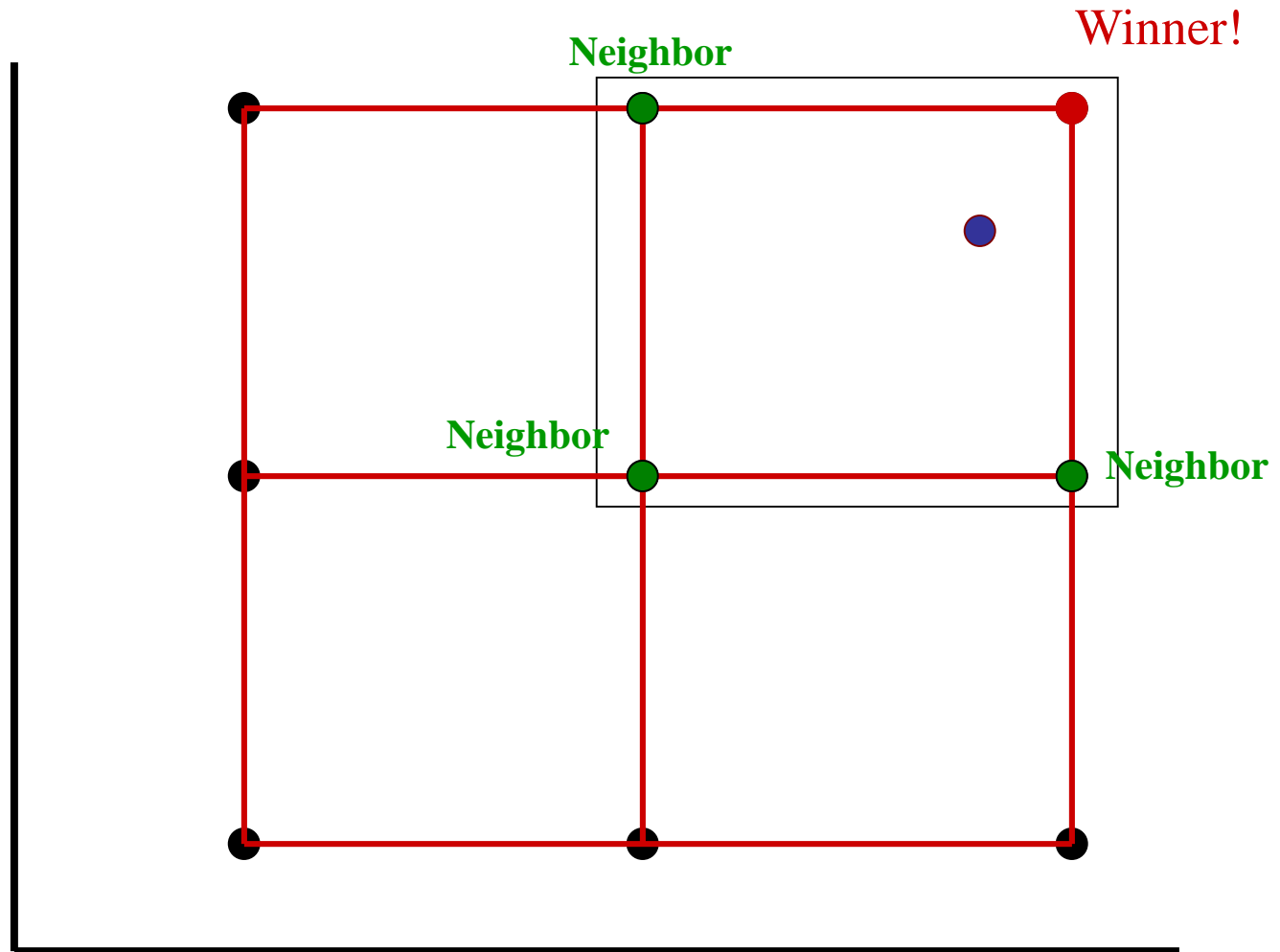


# El Análisis de Conglomerados con Redes Neuronales

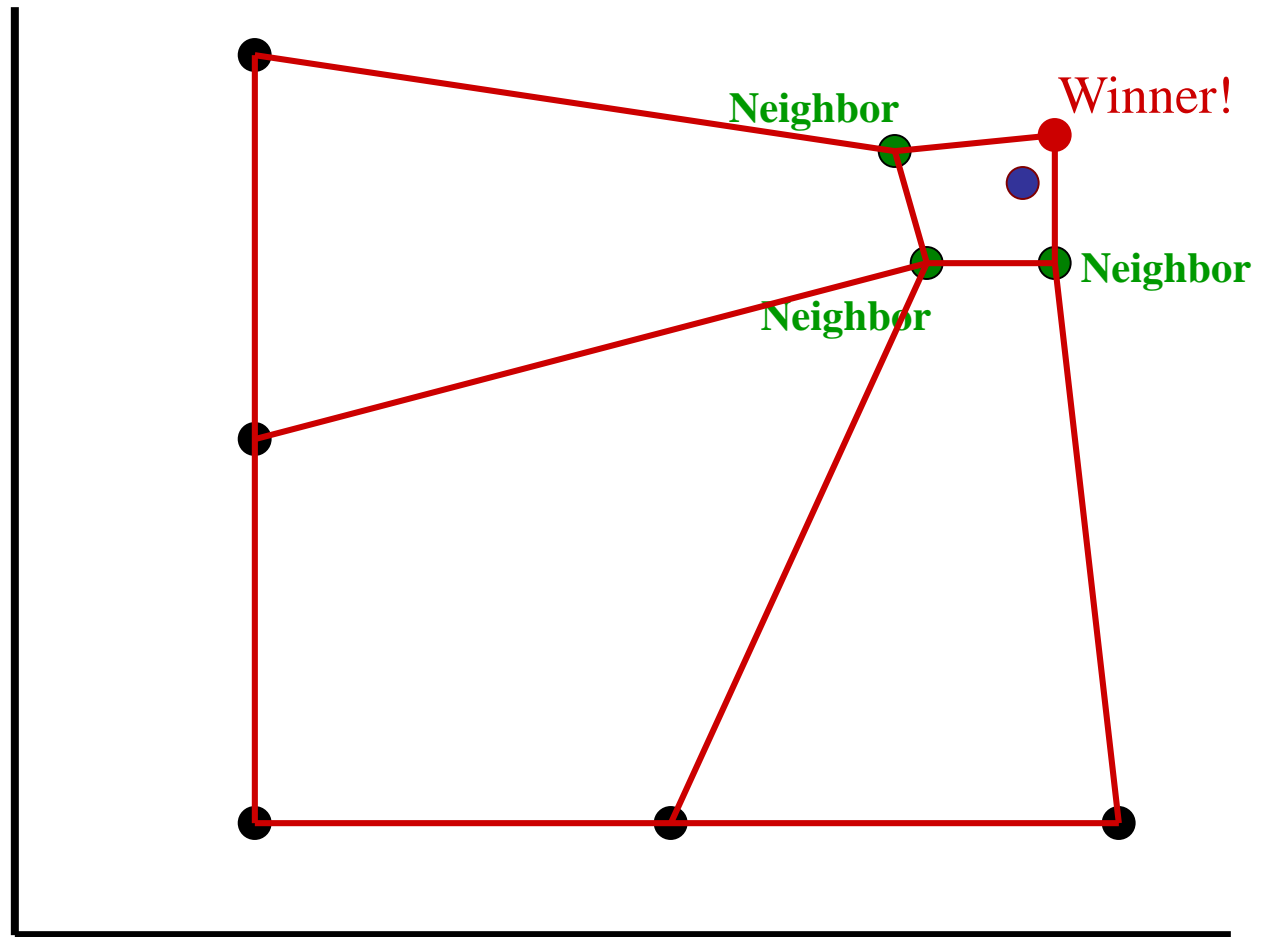
# Objetivos

- Discutir el concepto de SOM (self-organizing maps).
- Generar un SOM e interpretar los resultados.

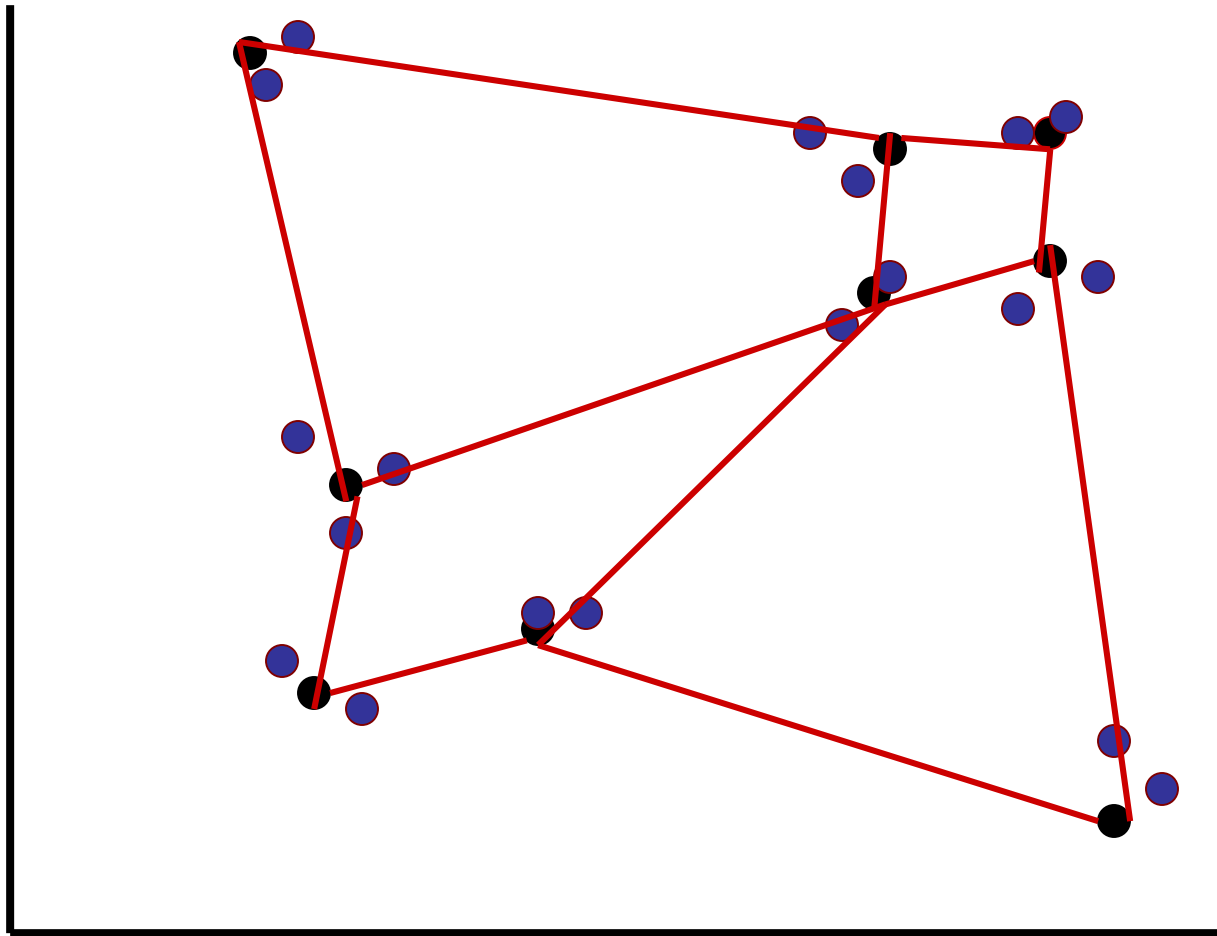
# Self-Organizing Maps



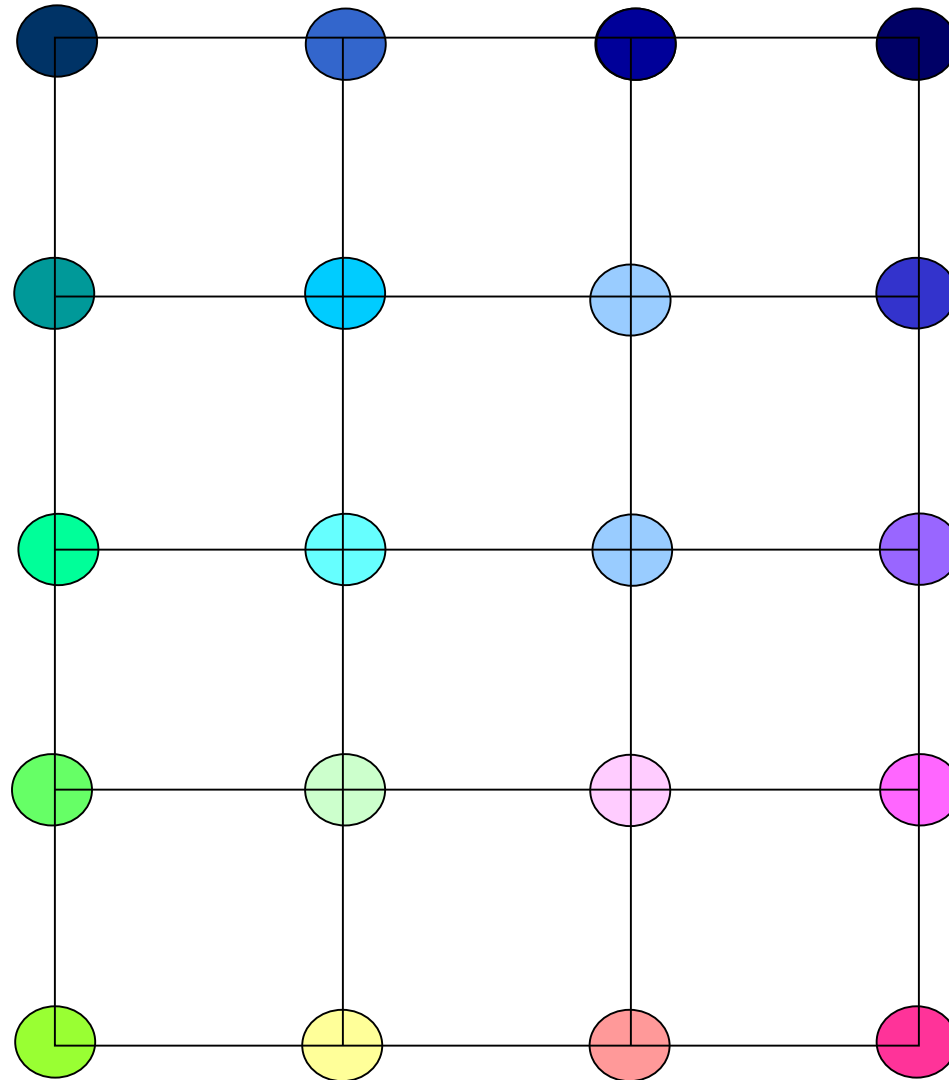
# Self-Organizing Maps



# Self-Organizing Maps



# Self-Organizing Maps



# Mapas Auto Organizables (SOM)

- Es una red neuronal que usa un algoritmo no supervisado de entrenamiento, y que a través de un proceso denominado auto organización, configura la salida de la red en una representación topológica de los datos originales.
- En contraste con las redes neuronales supervisadas, que requieren una variable objetivo en conjunción con las variables de entrada para encontrar un patrón, el SOM reduce los datos multidimensionales a un mapa bidimensional o red de neuronas.

# Razones para usar SOM

- Es un método numérico.
- Es una técnica no paramétrica.
- No se necesitan supuesto a priori acerca de la distribución de los datos.
- Es un algoritmo que puede detectar estructuras inesperadas o patrones por medio de un aprendizaje no supervisado,

# Algoritmo del SOM

Es modelo no lineal de variables latentes que genera una densidad de probabilidades en un espacio de datos multidimensional, usando un conjunto de variables latentes de menor dimensión. Este mapeo no lineal se describe a través de un modelo generalizado de regresión lineal:

$$y = W\phi(u)$$

# Algoritmo del SOM

Donde,

$u$  - Es un vector de variables latentes de  $L$  dimensiones.

$W$  - Es la matriz que genera el mapa explícito a partir del espacio latente a un espacio  $L$  dimensional múltiple grabado en los datos.

$\phi$  - Es un conjunto de  $R$  funciones, las cuales por lo general, se toman como Gaussianas.