



CAPACITACIÓN
PROFESIONAL

PEA Data Science

Aprendizaje No
Supervisado

MÓDULO 1

Docente: Marcos Santa Cruz
Vásquez

Reglas



Se requiere **puntualidad** para un mejor desarrollo del curso.



Para una mayor concentración **mantener silenciado el micrófono** durante la sesión.



Las preguntas se realizarán **a través del chat** y en caso de que lo requieran **podrán activar el micrófono**.



Realizar las actividades y/o tareas encomendadas en **los plazos determinados**.



Identificarse en la sala Zoom con el primer nombre y primer apellido.

ITINERARIO

Lunes y Viernes

*06:40 PM – 07:00 PM **Soporte técnico DMC***

*07:00 PM – 08:30 PM **Agenda***

*08:30 PM – 08:45 PM **Pausa Activa***

*08:45 PM – 10:00 PM **Agenda***

Horario de Atención Área Académica y Soporte

Lunes a Viernes 09:00 am a 10:30 pm / Sábado 09:00 am a 02:00pm

Modelos más comunes

Existen múltiples modelos, siendo los más comunes

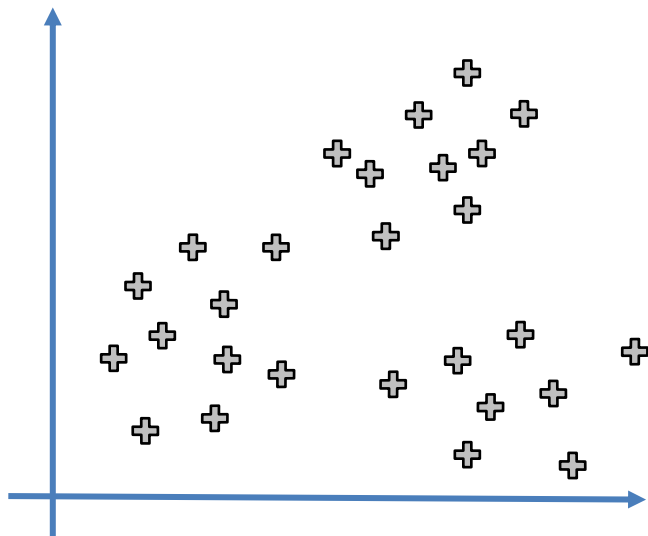
Aprendizaje Supervisado		Aprendizaje No supervisado		Aprendizaje Reforzado
Clasificación	Regresión	Clustering	Reducción dimensional	Q-learning
Logística	Regresión lineal	K-means	Pca	
Árbol de decisión - Bosque Rnd	Árbol de decisión - Bosque Rnd	Basado en la densidad	Análisis de discriminación lineal	

Idea de K-Means

Entendiendo K-Means

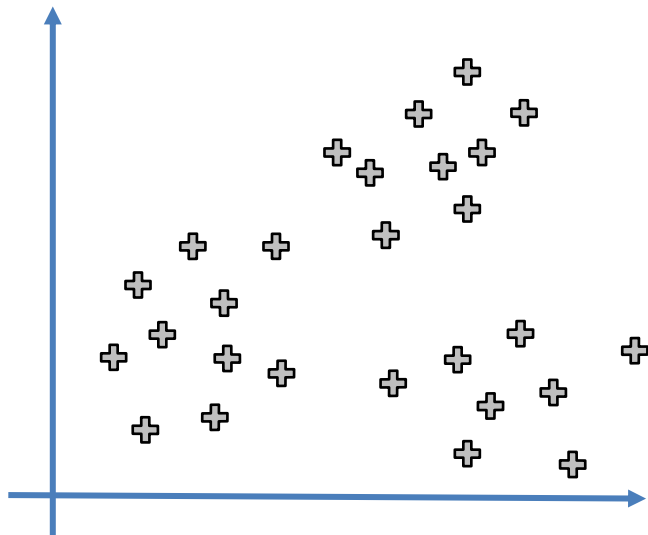
Para qué sirve K-Means

Antes de K-Means



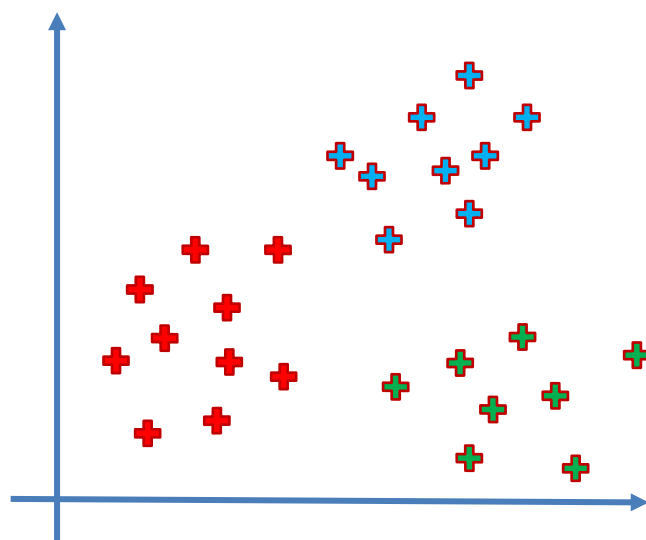
Para qué sirve K-Means

Antes de K-Means



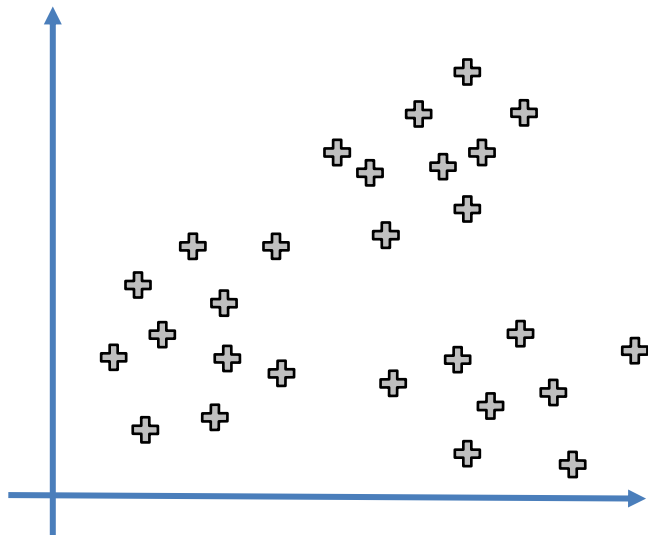
K-Means

Después de K-Means



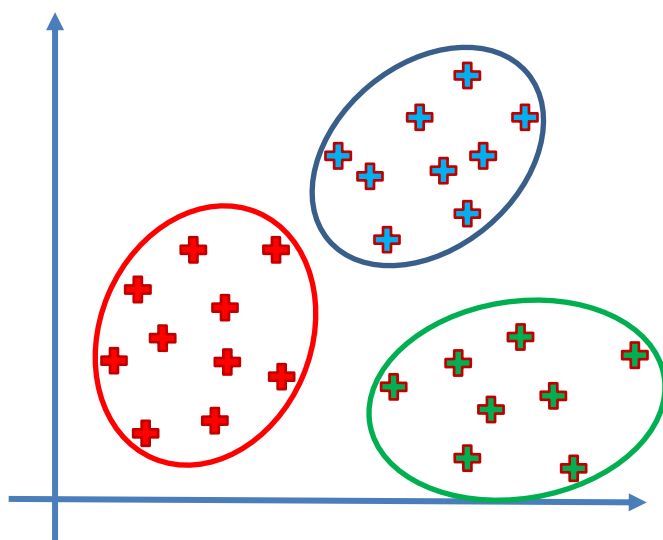
Para qué sirve K-Means

Antes de K-Means



K-Means

Después de K-Means



Cómo funciona

PASO 1: Elegir el número K de clusters



PASO 2: Seleccionar al azar K puntos, los baricentros (no necesariamente de nuestro dataset)



PASO 3: Asignar cada punto al baricentro más cercano → Esto formará los K clusters



PASO 4: Calcular y asignar el nuevo baricentro de cada cluster



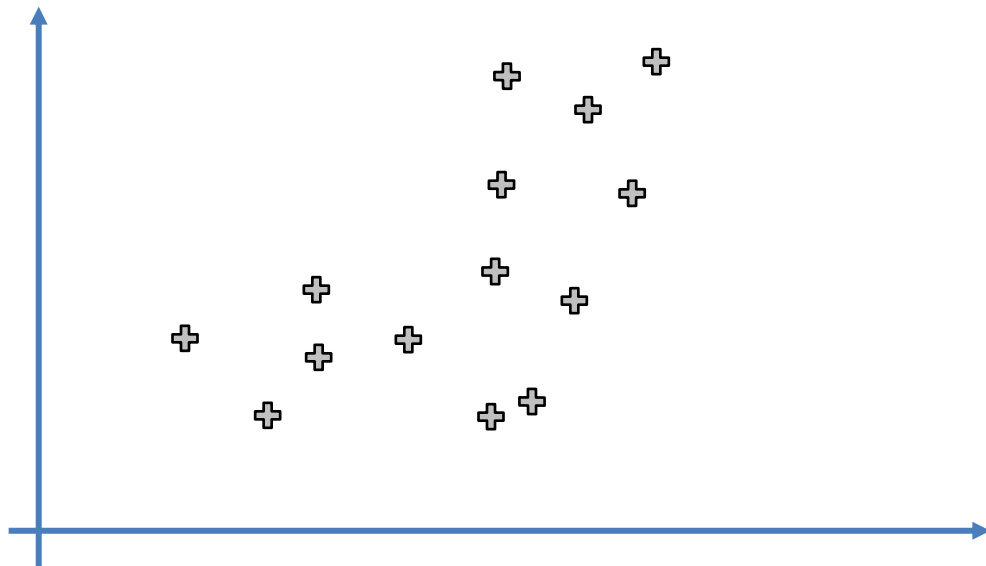
PASO 5: Reasignar cada punto de los datos a su baricentro más cercano.
Si ha habido nuevas asignaciones, ir al PASO 4, si no ir FIN.



El Modelo está Listo

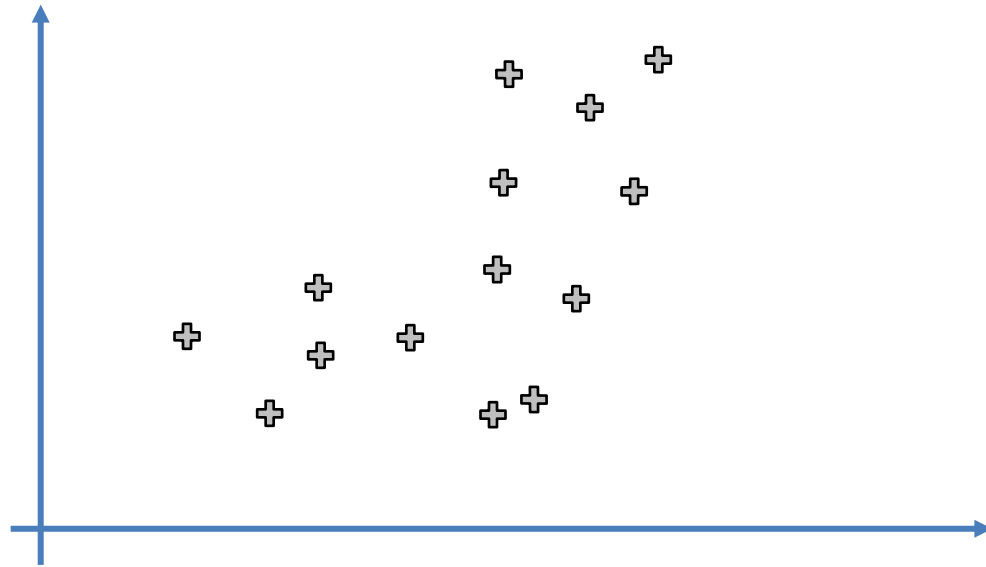
Algoritmo de K-Means

PASO 1: Elegir el número K de clusters: $K = 2$



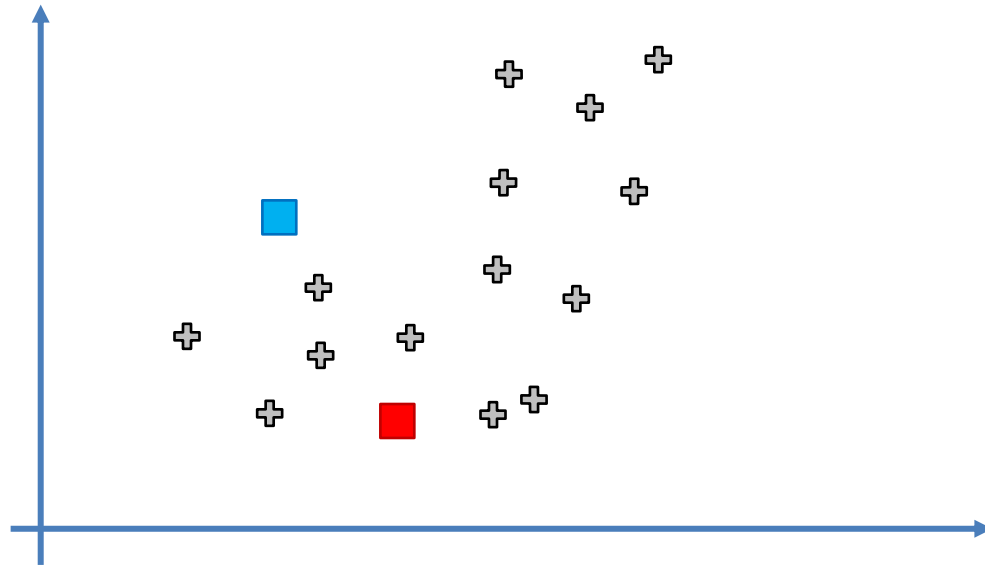
Algoritmo de K-Means

PASO 2: Seleccionar al azar K puntos, los baricentros (no necesariamente de nuestro dataset)



Algoritmo de K-Means

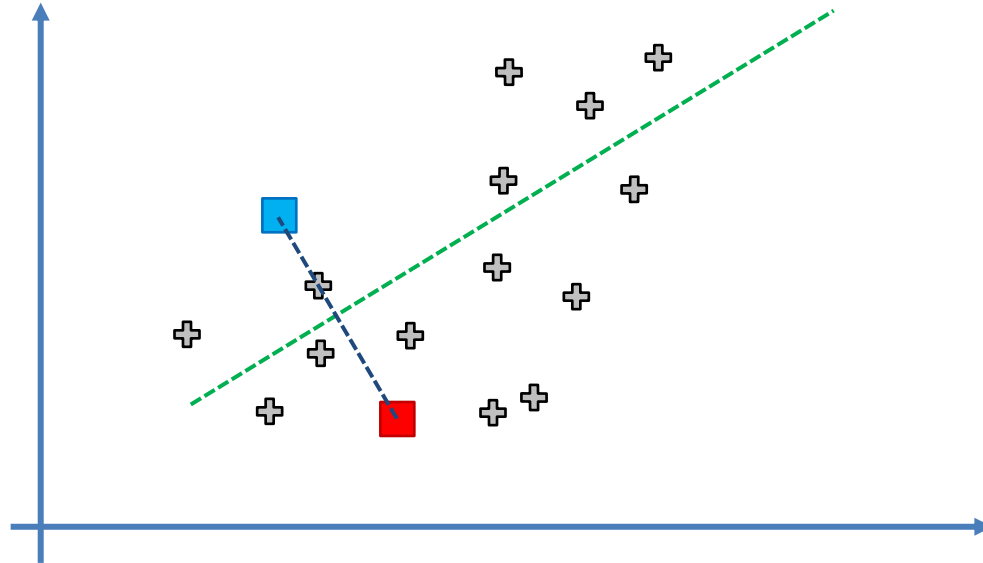
PASO 2: Seleccionar al azar K puntos, los baricentros (no necesariamente de nuestro dataset)



Algoritmo de K-Means

PASO 3: Asignar cada punto al baricentro más cercano
clusters →

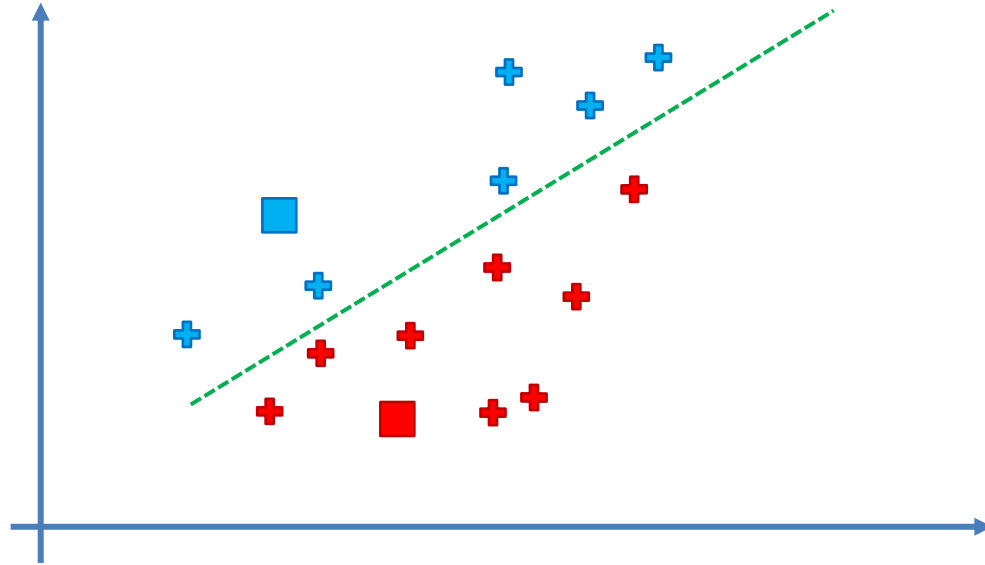
Esto formará los K



Algoritmo de K-Means

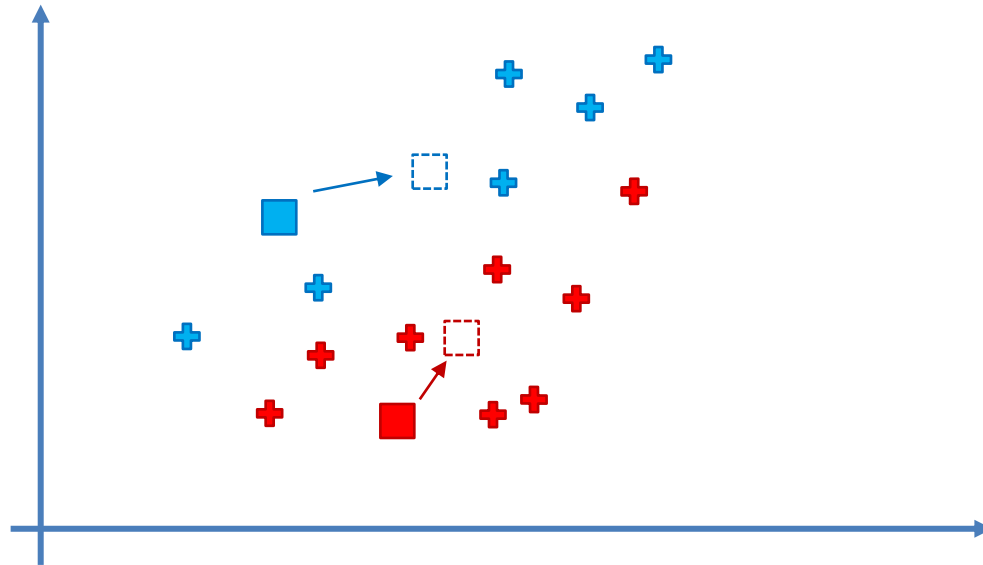
PASO 3: Asignar cada punto al baricentro más cercano → clusters

Esto formará los K



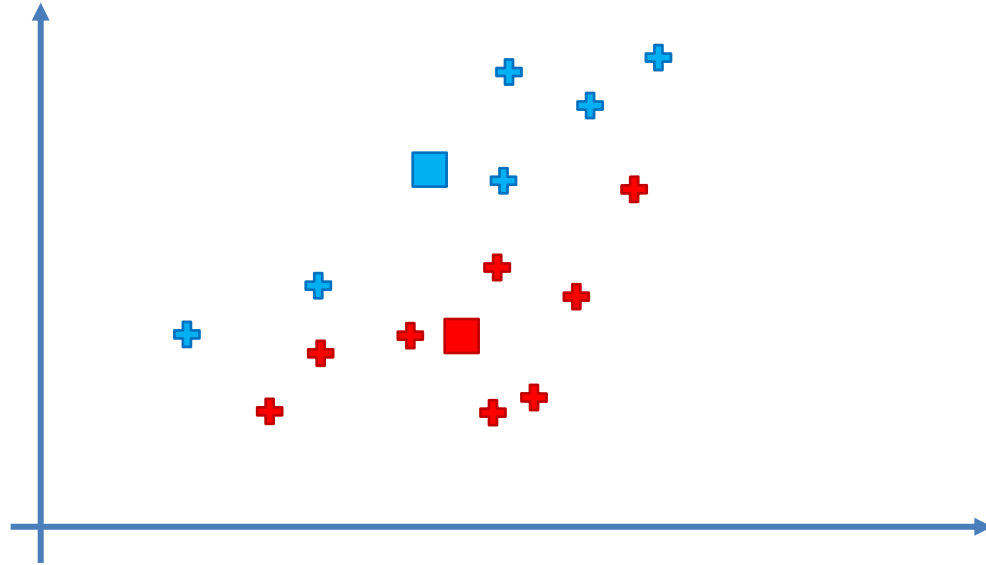
Algoritmo de K-Means

PASO 4: Calcular y asignar el nuevo baricentro de cada cluster



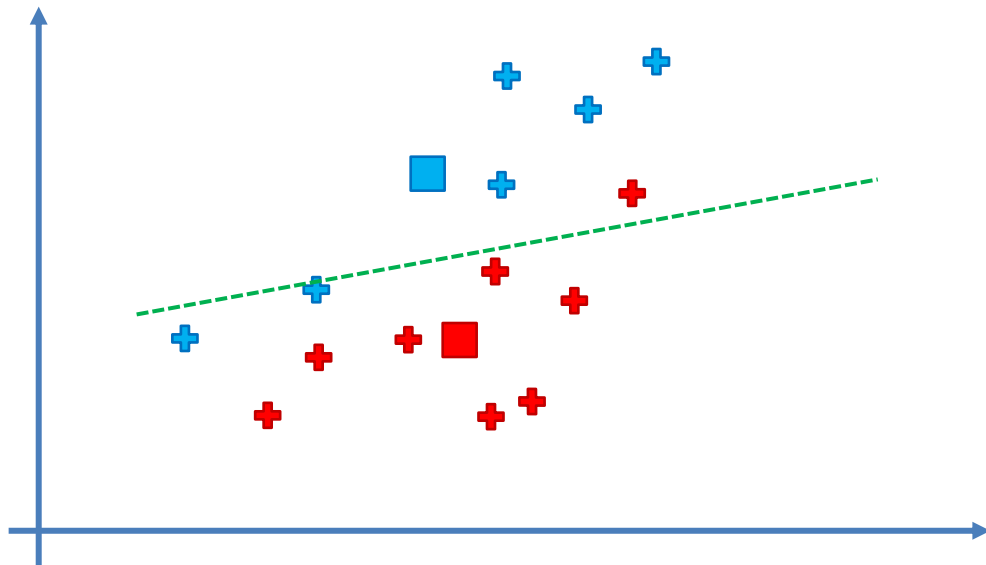
Algoritmo de K-Means

PASO 4: Calcular y asignar el nuevo baricentro de cada cluster



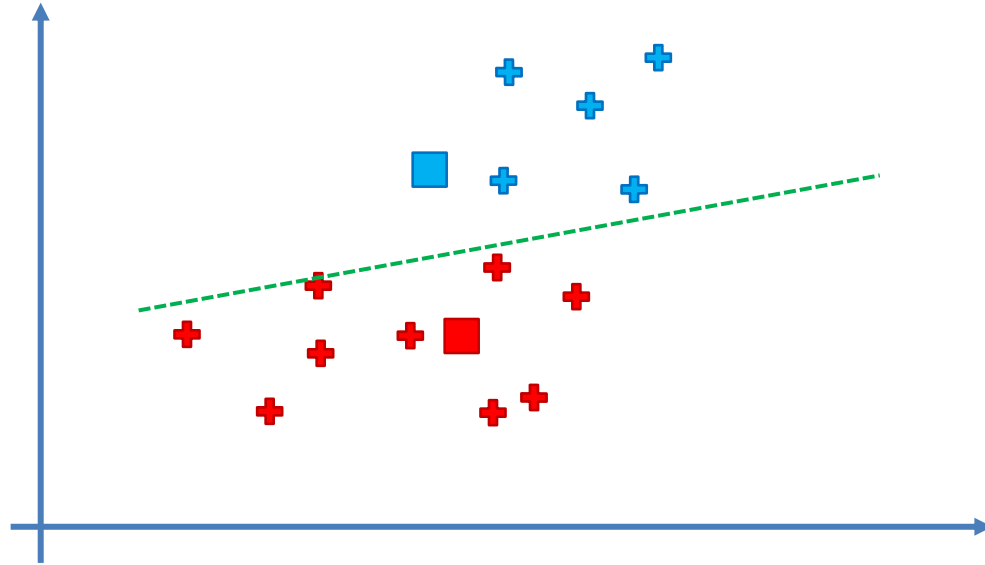
Algoritmo de K-Means

PASO 5: Reasignar cada punto de los datos a su baricentro más cercano.
Si ha habido nuevas asignaciones, ir al PASO 4, si no ir FIN.



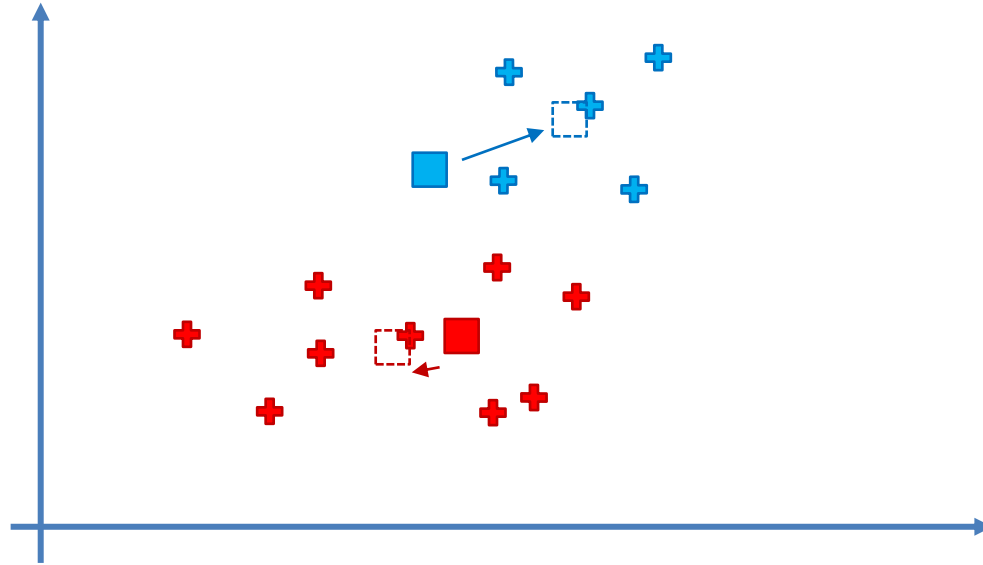
Algoritmo de K-Means

PASO 5: Reasignar cada punto de los datos a su baricentro más cercano.
Si ha habido nuevas asignaciones, ir al PASO 4, si no ir FIN.



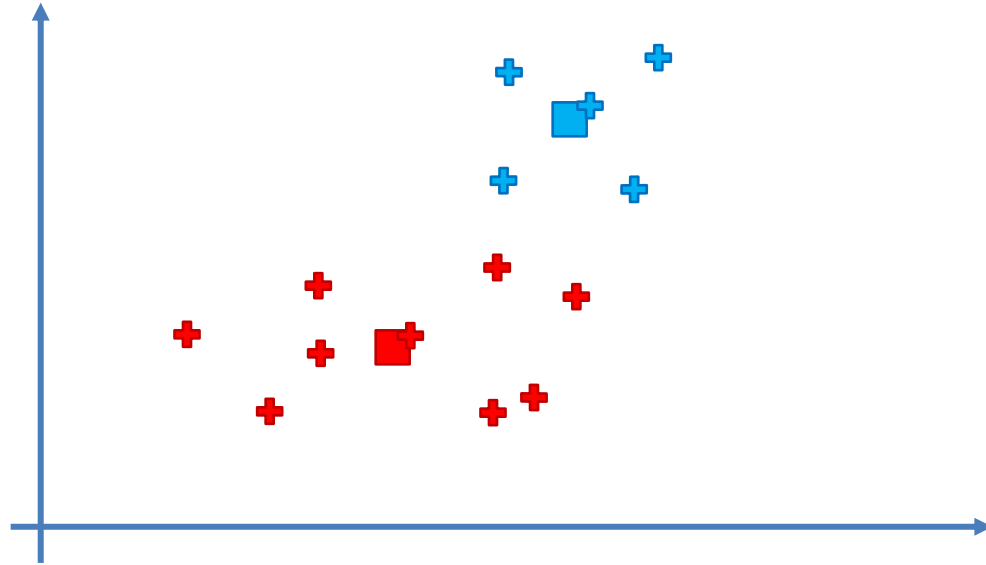
Algoritmo de K-Means

PASO 4: Calcular y asignar el nuevo baricentro de cada cluster



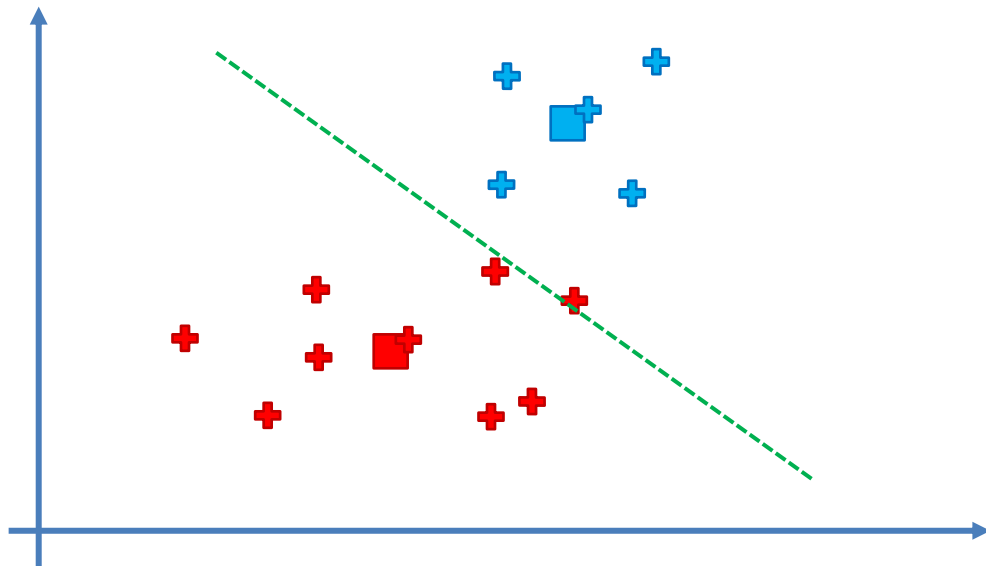
Algoritmo de K-Means

PASO 4: Calcular y asignar el nuevo baricentro de cada cluster



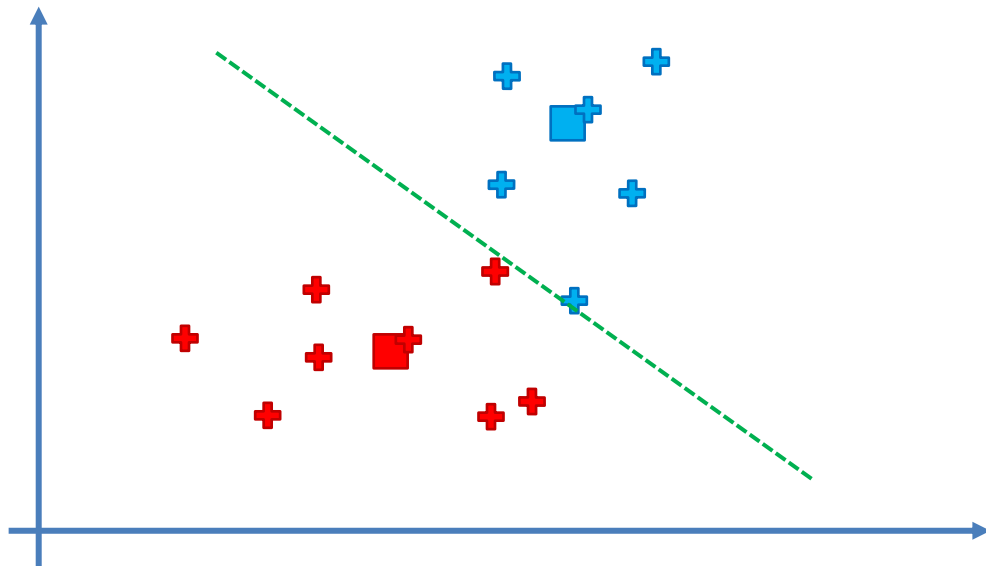
Algoritmo de K-Means

PASO 5: Reasignar cada punto de los datos a su baricentro más cercano.
Si ha habido nuevas asignaciones, ir al PASO 4, si no ir FIN.



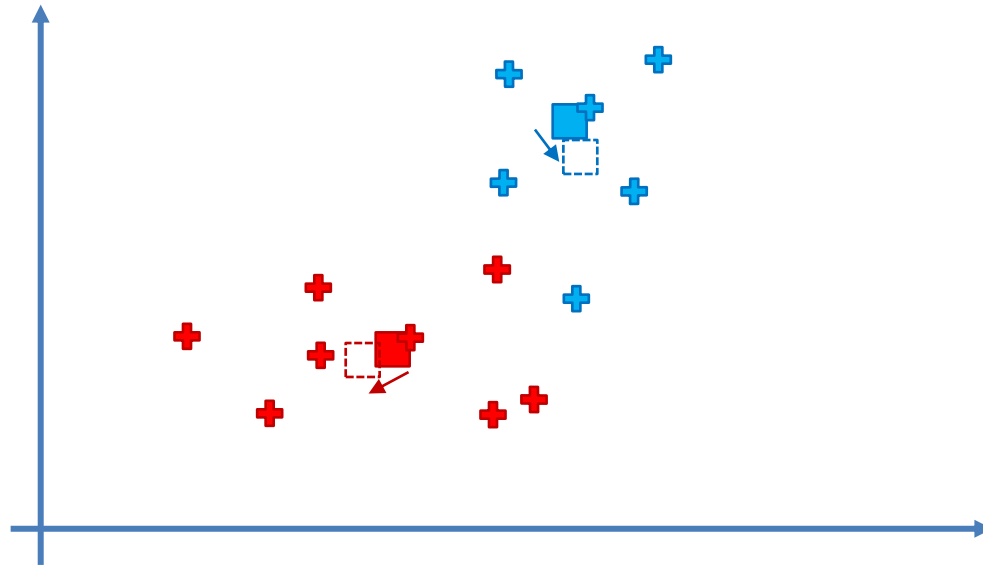
Algoritmo de K-Means

PASO 5: Reasignar cada punto de los datos a su baricentro más cercano.
Si ha habido nuevas asignaciones, ir al PASO 4, si no ir FIN.



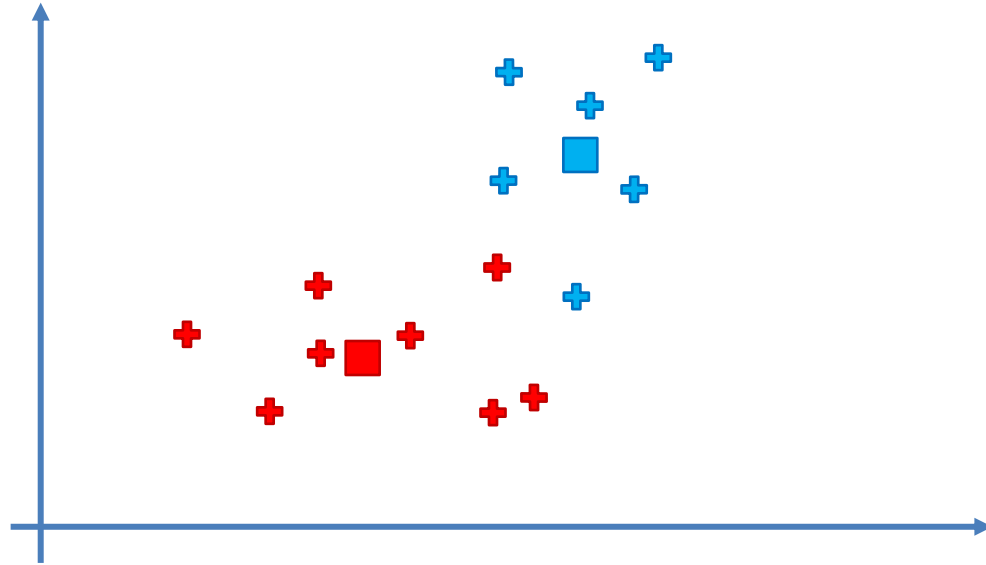
Algoritmo de K-Means

PASO 4: Calcular y asignar el nuevo baricentro de cada cluster



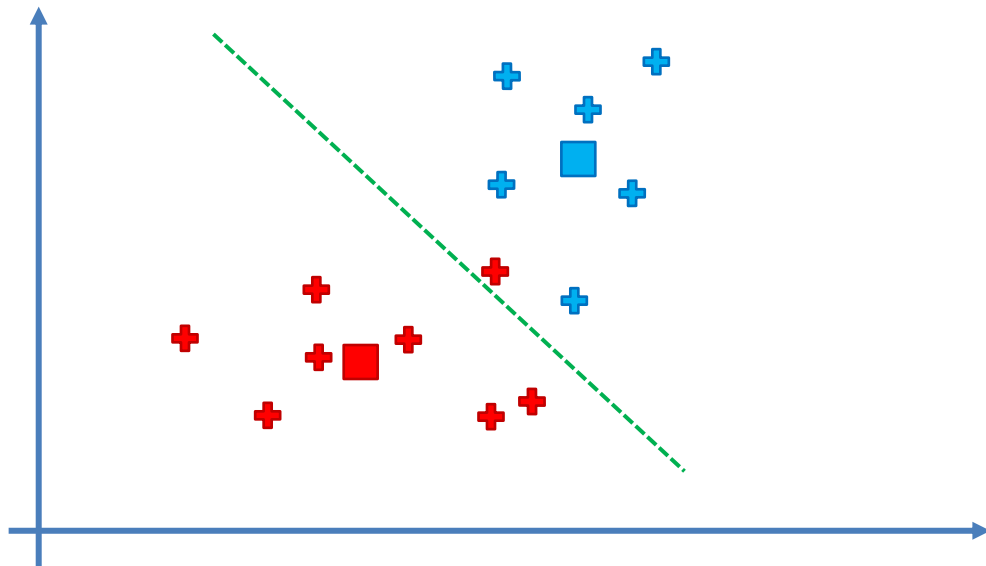
Algoritmo de K-Means

PASO 4: Calcular y asignar el nuevo baricentro de cada cluster



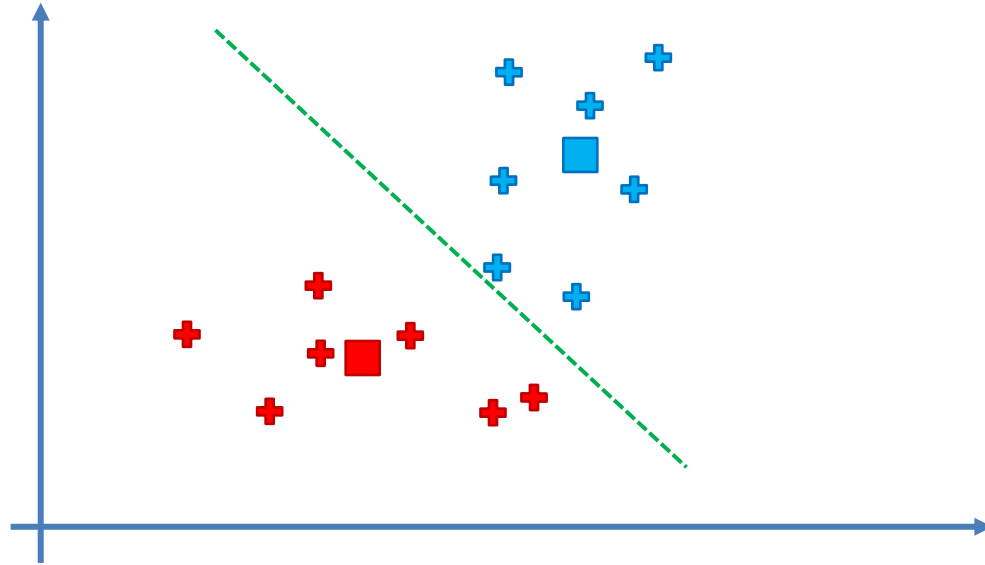
Algoritmo de K-Means

PASO 5: Reasignar cada punto de los datos a su baricentro más cercano.
Si ha habido nuevas asignaciones, ir al PASO 4, si no ir FIN.



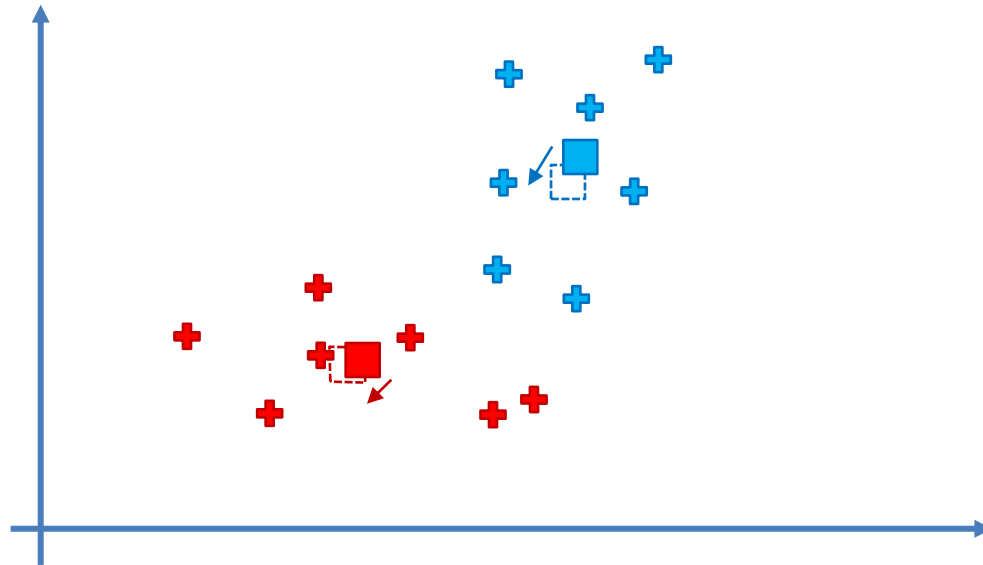
Algoritmo de K-Means

PASO 5: Reasignar cada punto de los datos a su baricentro más cercano.
Si ha habido nuevas asignaciones, ir al PASO 4, si no ir FIN.



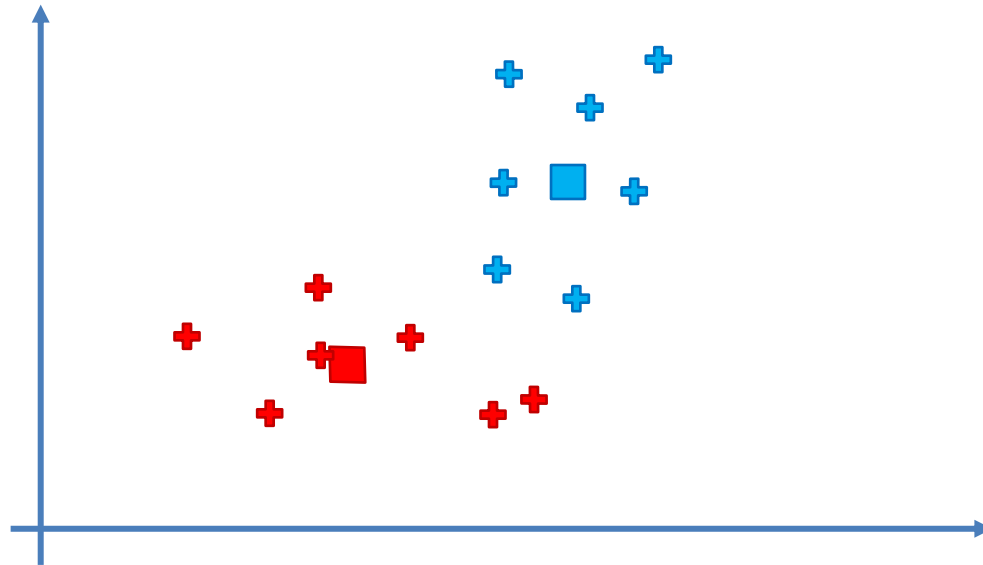
Algoritmo de K-Means

PASO 4: Calcular y asignar el nuevo baricentro de cada cluster



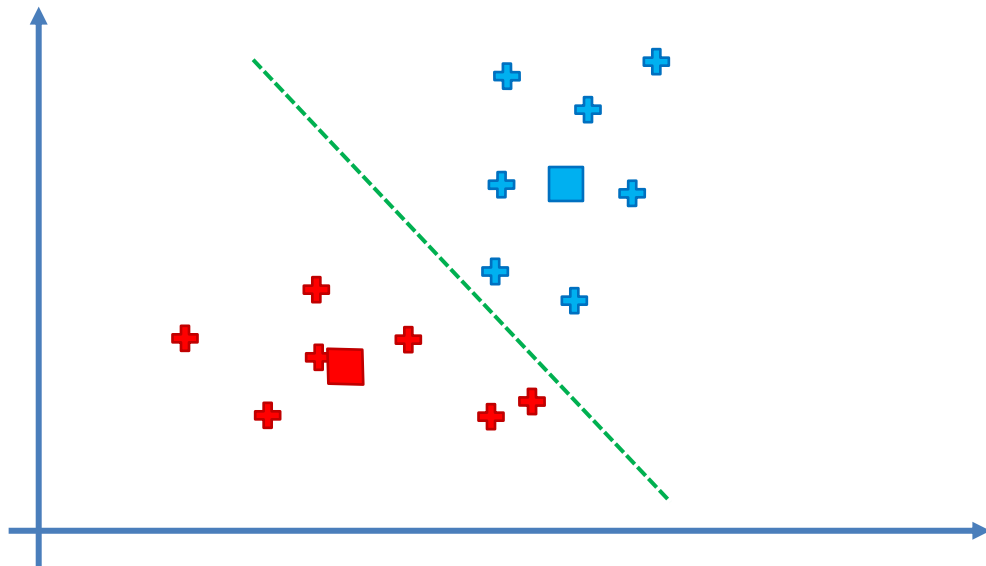
Algoritmo de K-Means

PASO 4: Calcular y asignar el nuevo baricentro de cada cluster



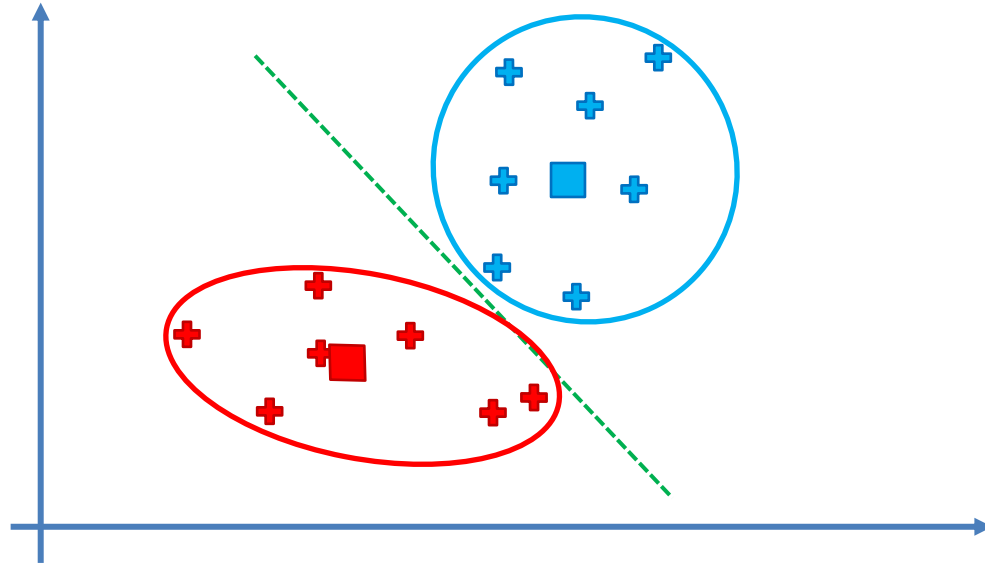
Algoritmo de K-Means

PASO 5: Reasignar cada punto de los datos a su baricentro más cercano.
Si ha habido nuevas asignaciones, ir al PASO 4, si no ir FIN.



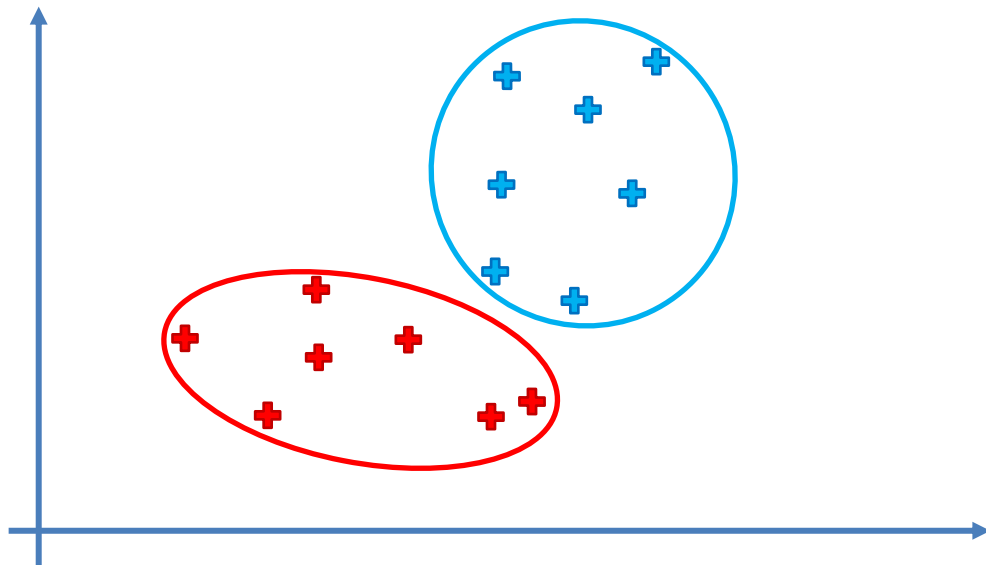
Algoritmo de K-Means

FIN: El modelo está listo



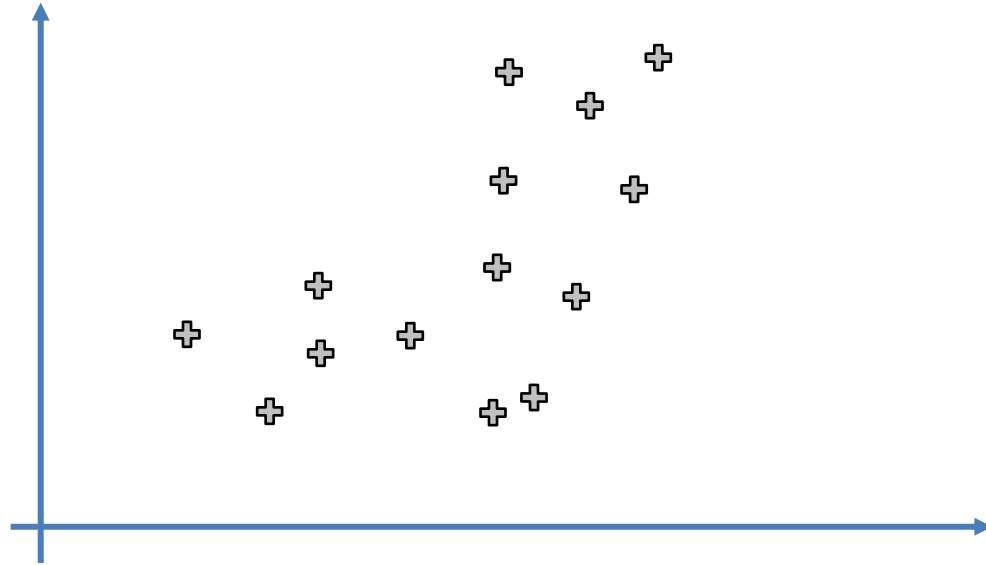
Algoritmo de K-Means

FIN: El modelo está listo



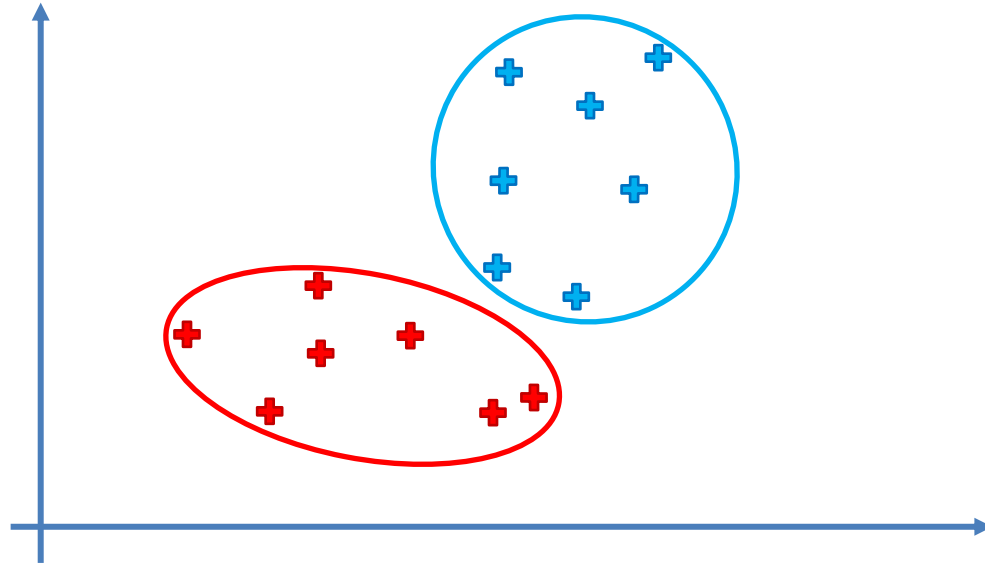
Algoritmo de K-Means

PASO 2: Seleccionar al azar K puntos, los baricentros (no necesariamente de nuestro dataset)



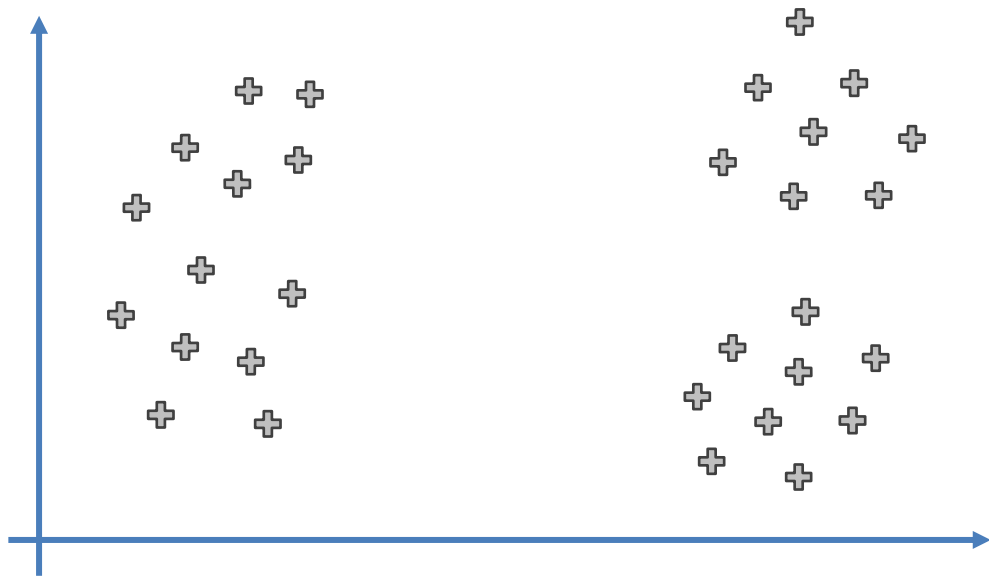
Algoritmo de K-Means

FIN: Your Model Is Ready

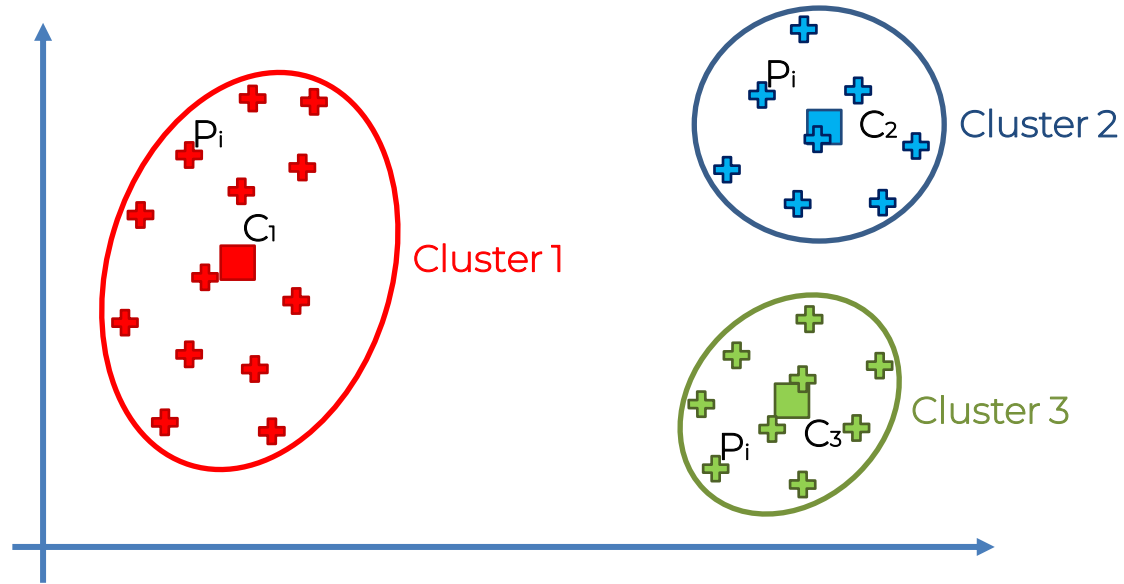


Idea de K-Means Intuition: Elegir el número correcto de clusters

Elegir el número correcto de clusters



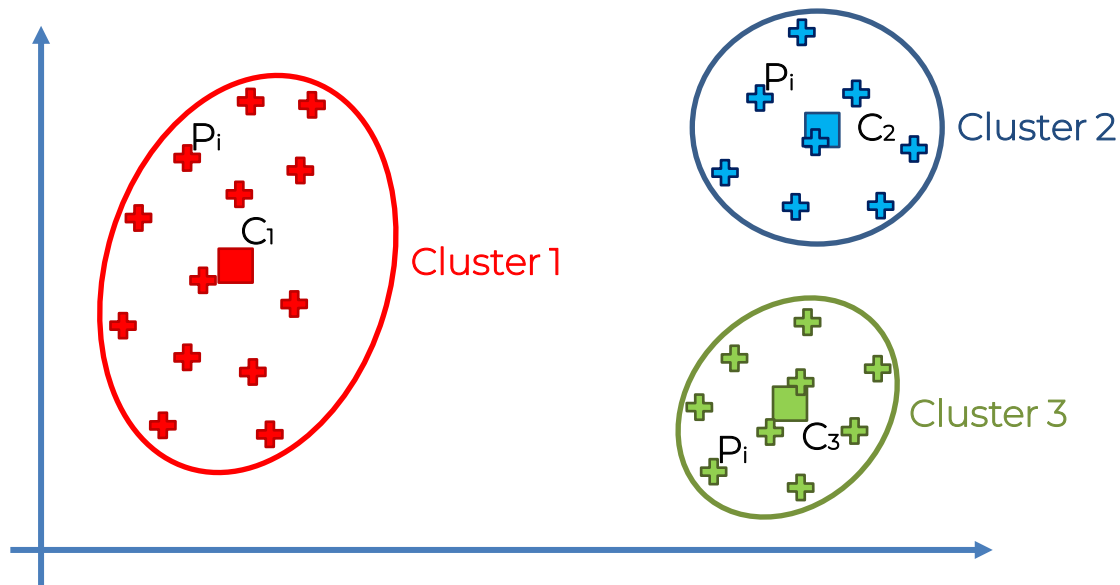
Elegir el número correcto de clusters



Elegir el número correcto de clusters

$$WCSS = \sum_{P_i \in \text{Cluster } 1} d(P_i, C_1)^2 + \sum_{P_i \in \text{Cluster } 2} d(P_i, C_2)^2 + \sum_{P_i \in \text{Cluster } 3} d(P_i, C_3)^2$$

Elegir el número correcto de clusters

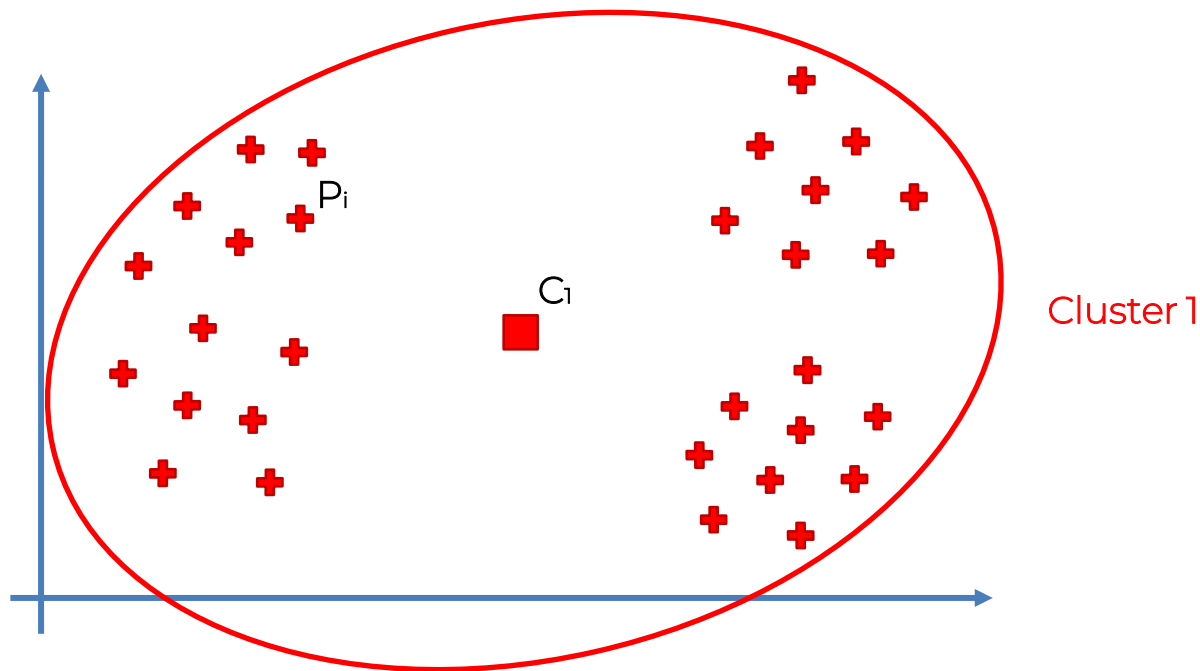


$$WCSS = \sum_{P_i \in \text{Cluster 1}} d(P_i, C_1)^2 + \sum_{P_i \in \text{Cluster 2}} d(P_i, C_2)^2 + \sum_{P_i \in \text{Cluster 3}} d(P_i, C_3)^2$$

Elegir el número correcto de clusters

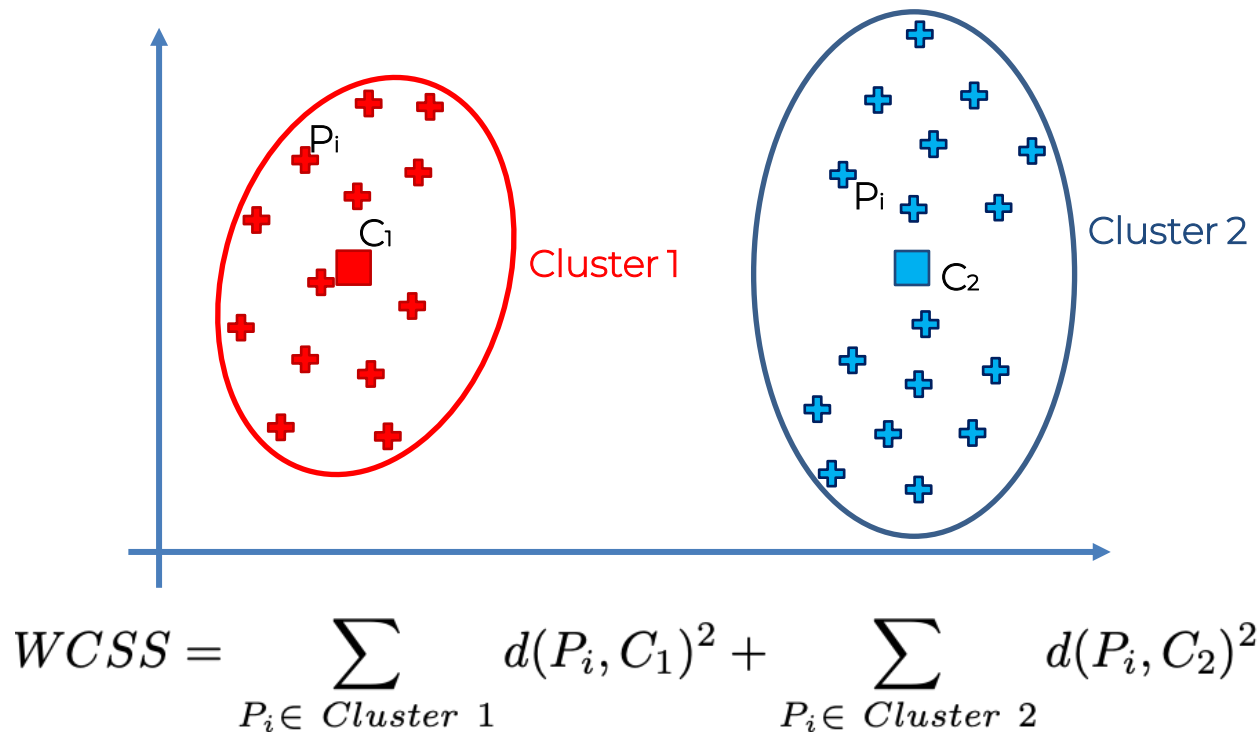
Rebobinemos...

Elegir el número correcto de clusters

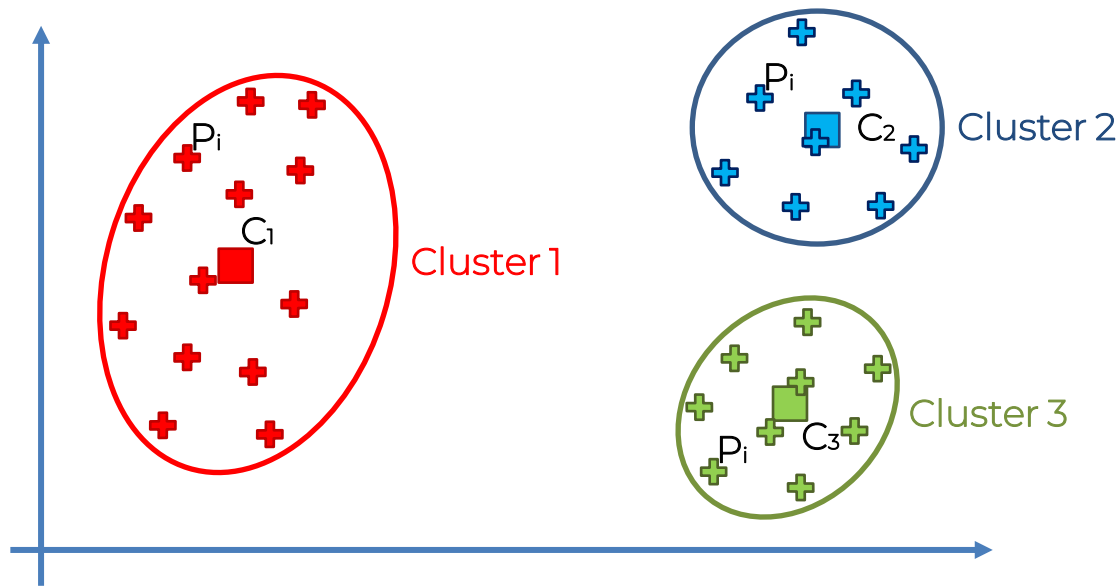


$$WCSS = \sum_{P_i \in \text{Cluster } 1} d(P_i, C_1)^2$$

Elegir el número correcto de clusters

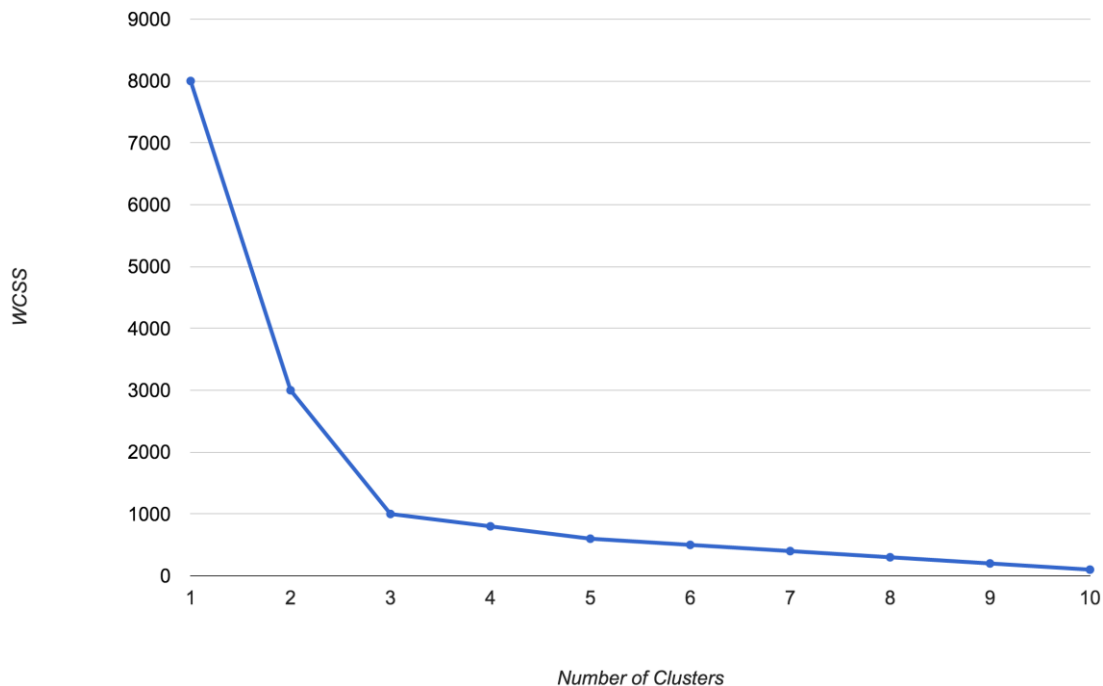


Elegir el número correcto de clusters



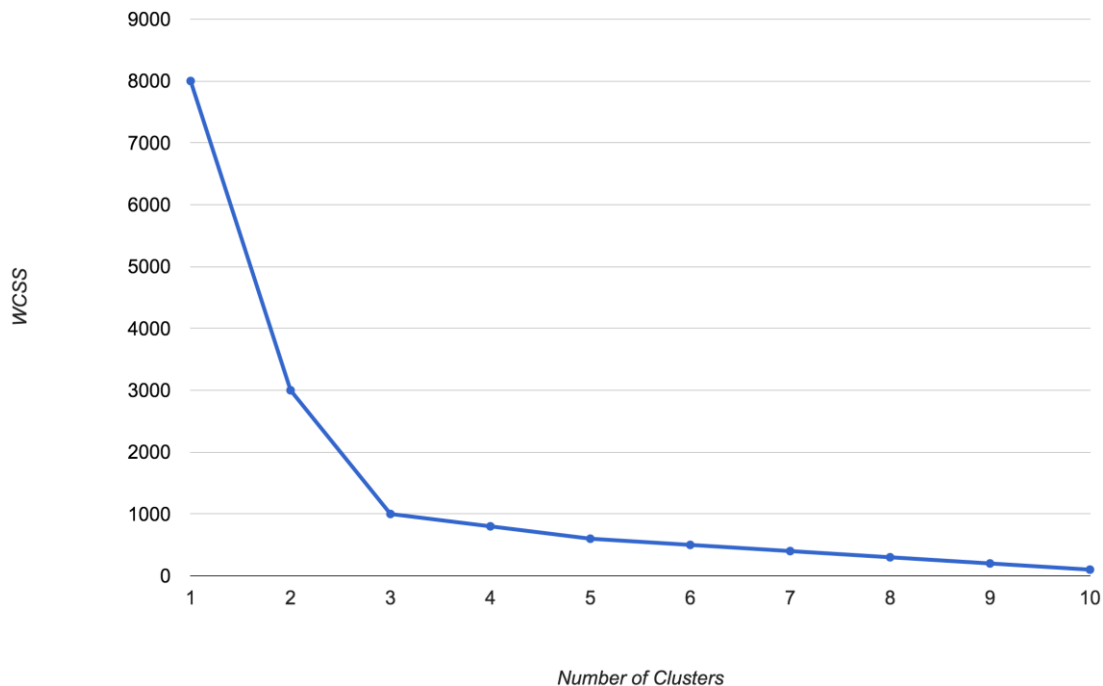
$$WCSS = \sum_{P_i \in \text{Cluster 1}} d(P_i, C_1)^2 + \sum_{P_i \in \text{Cluster 2}} d(P_i, C_2)^2 + \sum_{P_i \in \text{Cluster 3}} d(P_i, C_3)^2$$

Elegir el número correcto de clusters



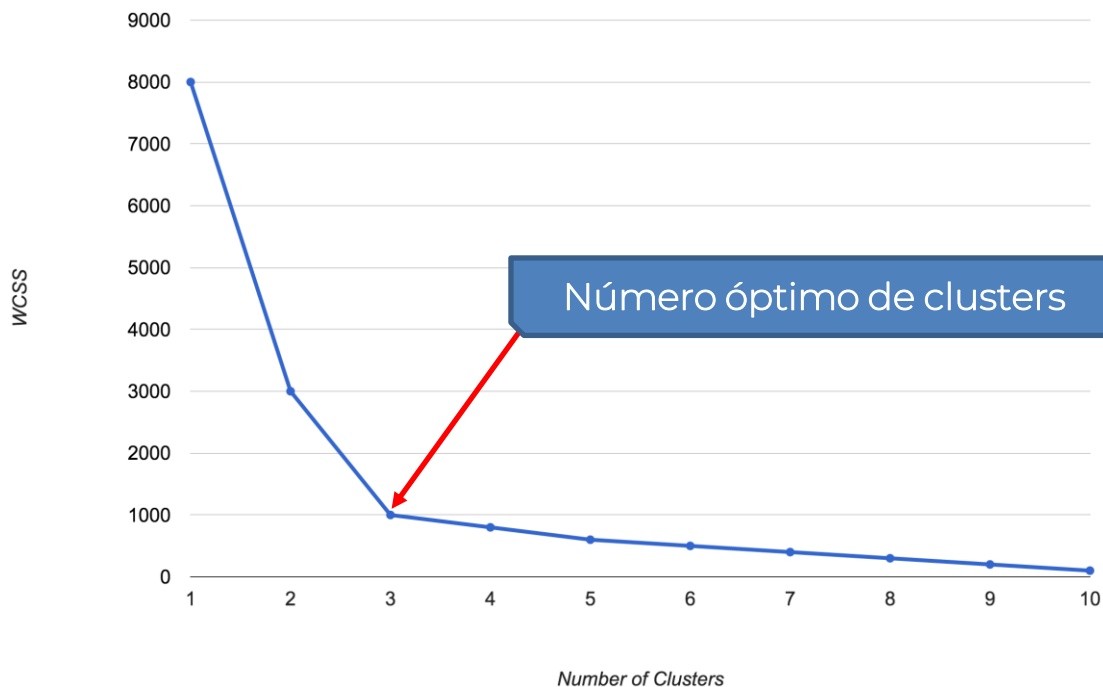
Elegir el número correcto de clusters

La técnica del codo



Elegir el número correcto de clusters

La técnica del codo

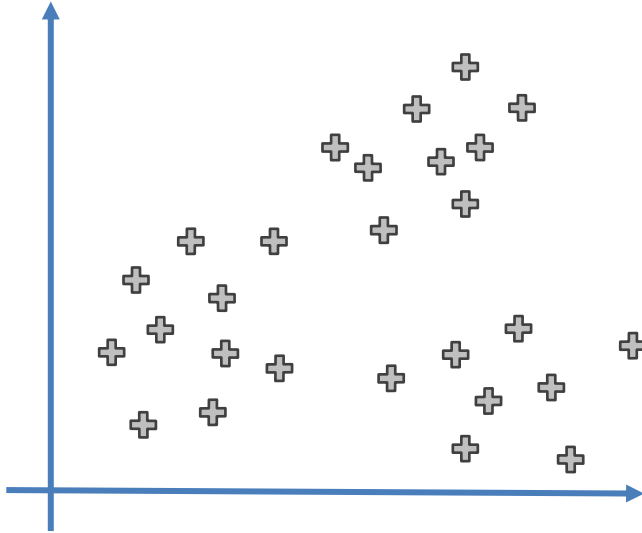


Idea del Clustering Jerárquico: Cómo funciona

Cómo funciona el Clustering

Jerárquico

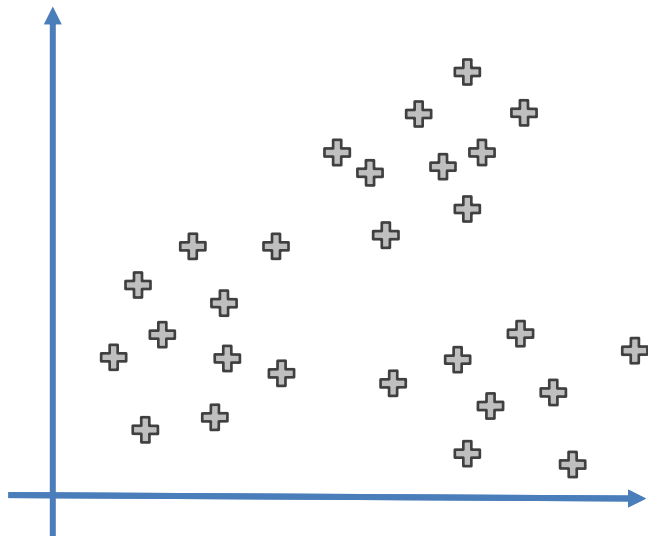
Antes del CJ



Cómo funciona el Clustering

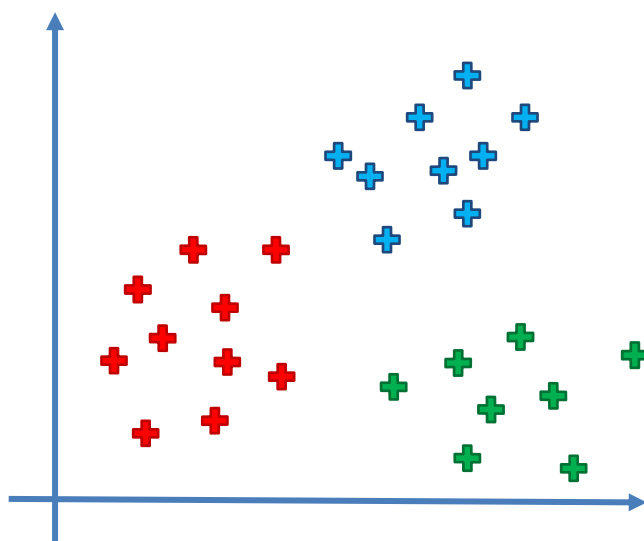
Jerárquico

Antes del CJ



CJ

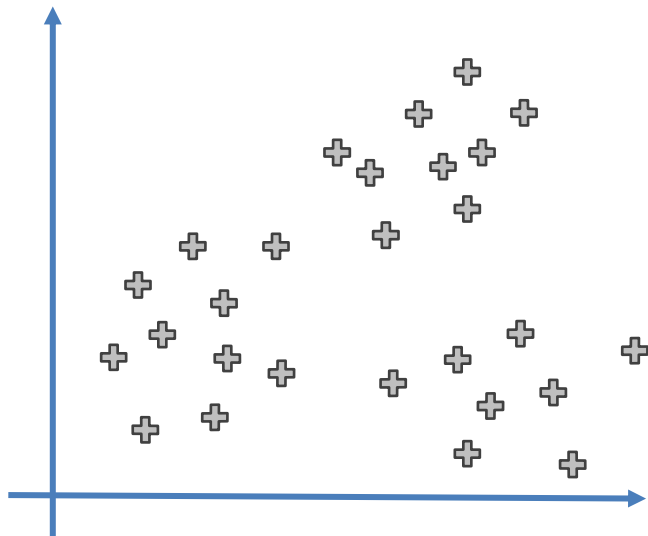
Después del CJ



Cómo funciona el Clustering

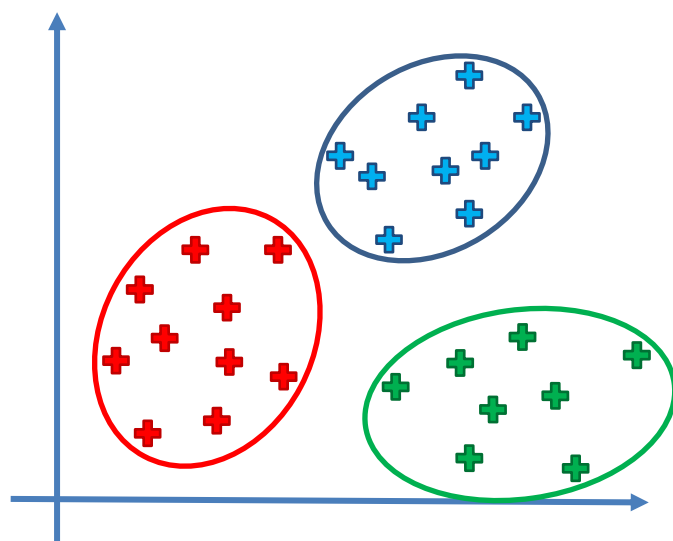
Jerárquico

Antes del CJ



CJ

Después del CJ



Igual que K-Means pero con diferente
proceso

NOTA:
Aglomerativo
&
Divisitivo



Clustering Jerárquico Aglomerativo

PASO 1: Hacer que cada punto sea un propio cluster. ➡ Así tendremos N clusters



PASO 2: Elegir los dos puntos más cercanos y juntarlos en un único cluster ➡ $N-1$ clusters



PASO 3: Elegir los dos clusters más cercanos y juntarlos en un único cluster ➡ $N-2$ clusters

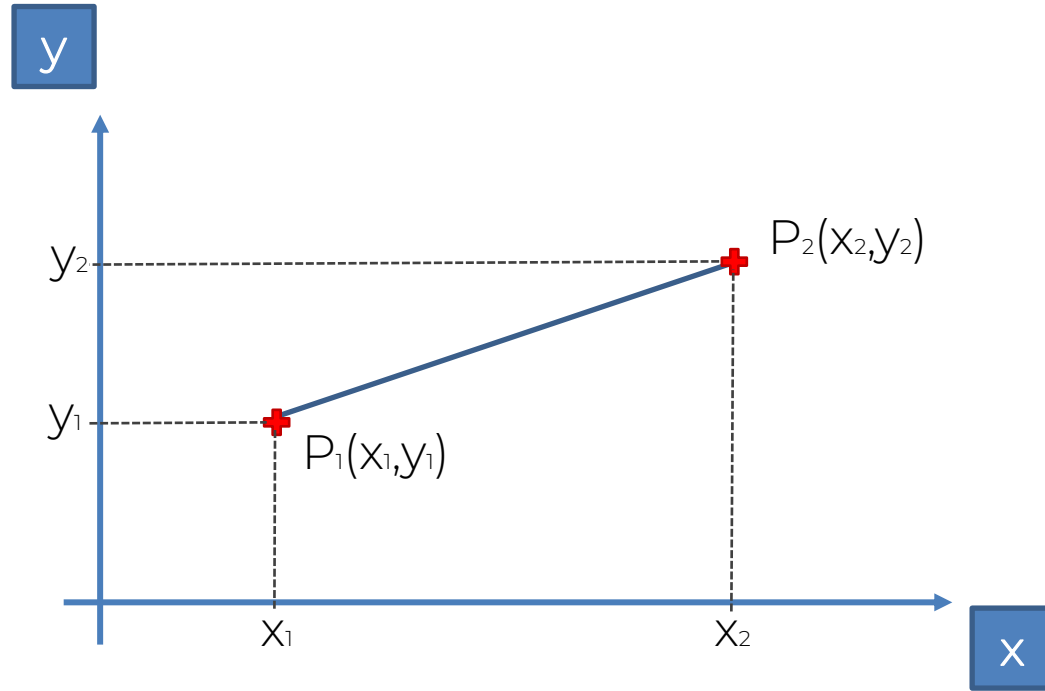


PASO 4: Repetir el PASO 3 hasta solo tener un único cluster



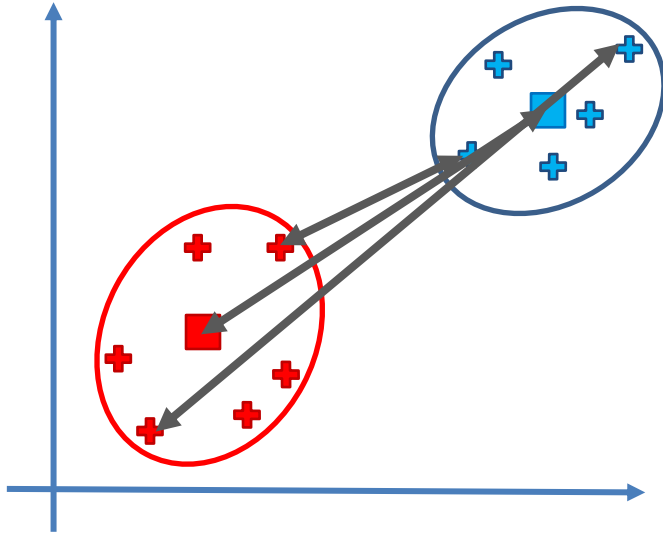
FIN

Distancia Euclídea



$$\text{Euclidean Distance between } P_1 \text{ and } P_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Distancia entre Clusters

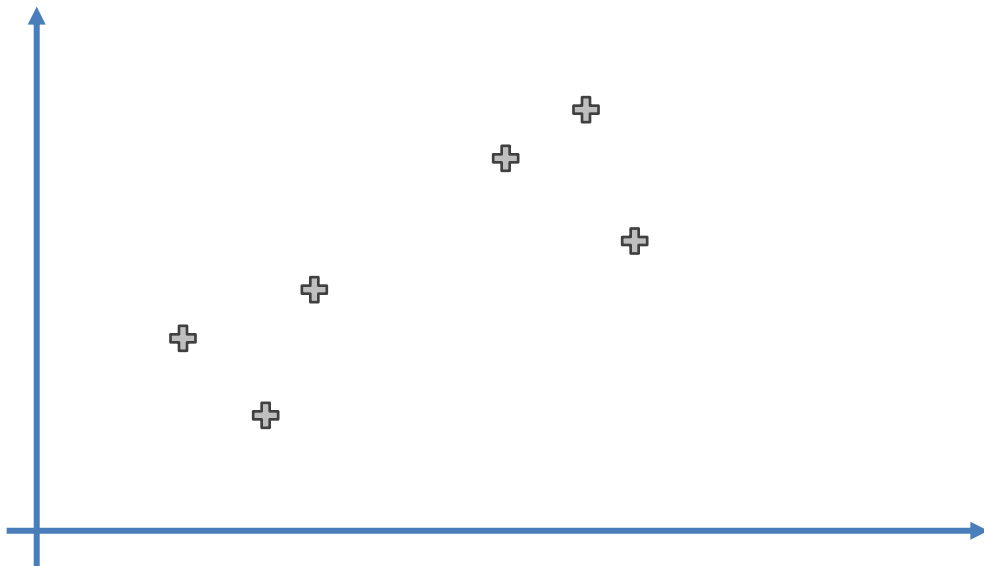


Distancia entre dos Clusters:

- Opción 1: Puntos más cercanos
- Opción 2: Puntos más alejados
- Opción 3: Distancia media
- Opción 4: Distancia entre sus baricentros

Clustering Jerárquico Aglomerativo

Consideremos el siguiente data set de $N = 6$ puntos

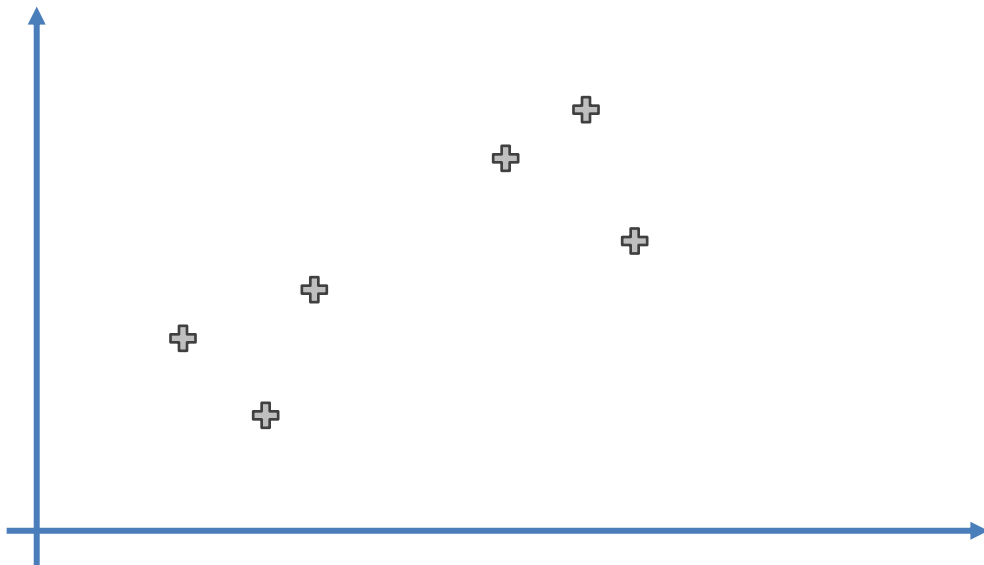


Clustering Jerárquico Aglomerativo

PASO 1: Hacer que cada punto sea un propio cluster.
clusters



Así tendremos 6

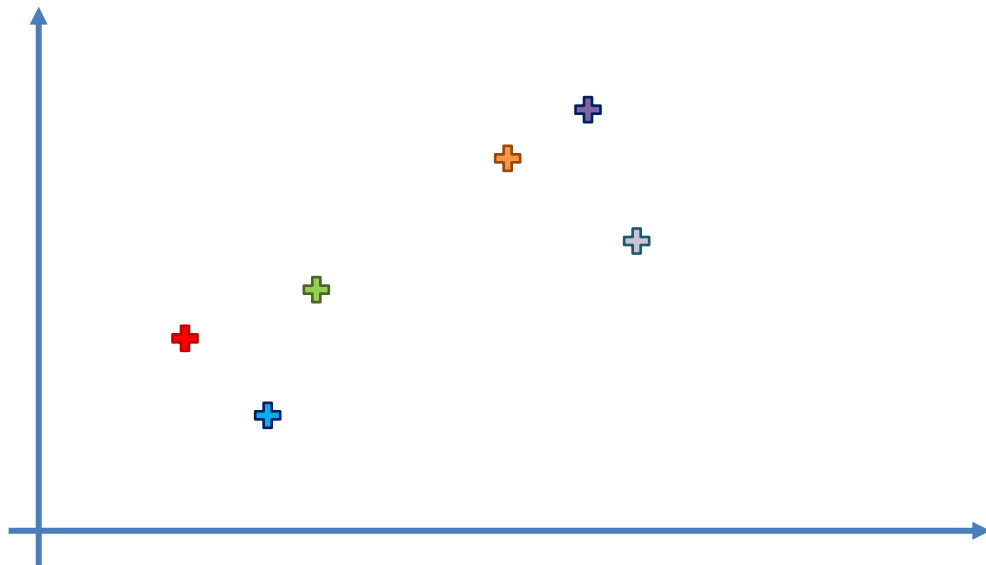


Clustering Jerárquico Aglomerativo

PASO 1: Hacer que cada punto sea un propio cluster.
clusters



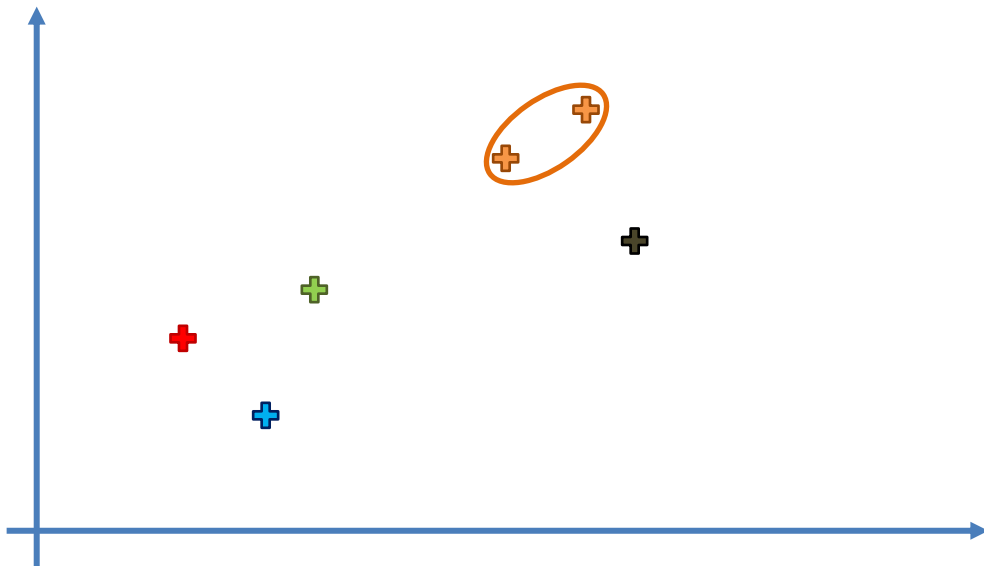
Así tendremos 6



Clustering Jerárquico Aglomerativo

PASO 2: Elegir los dos puntos más cercanos y juntarlos en un único cluster

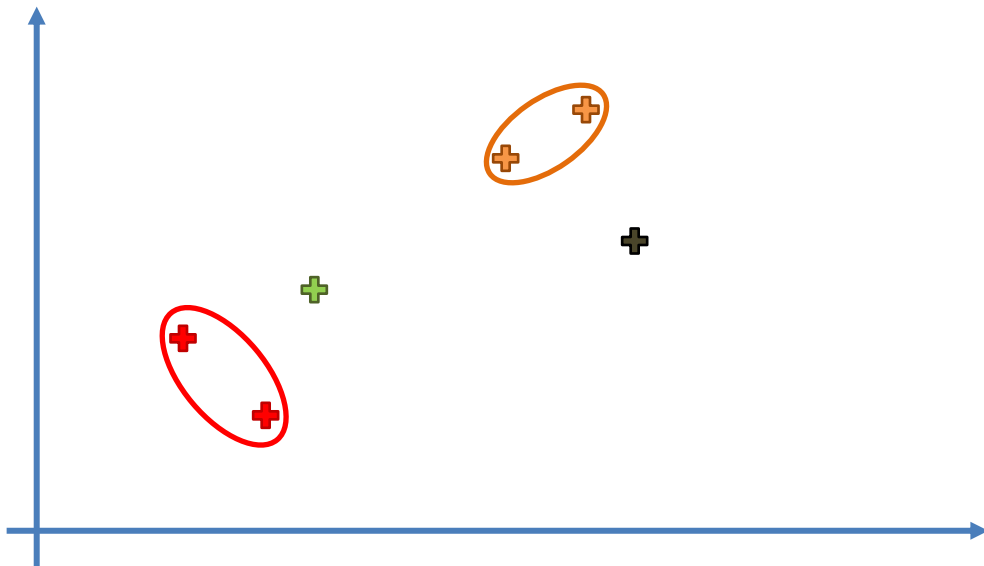
→ Así nos quedan 5 clusters



Clustering Jerárquico Aglomerativo

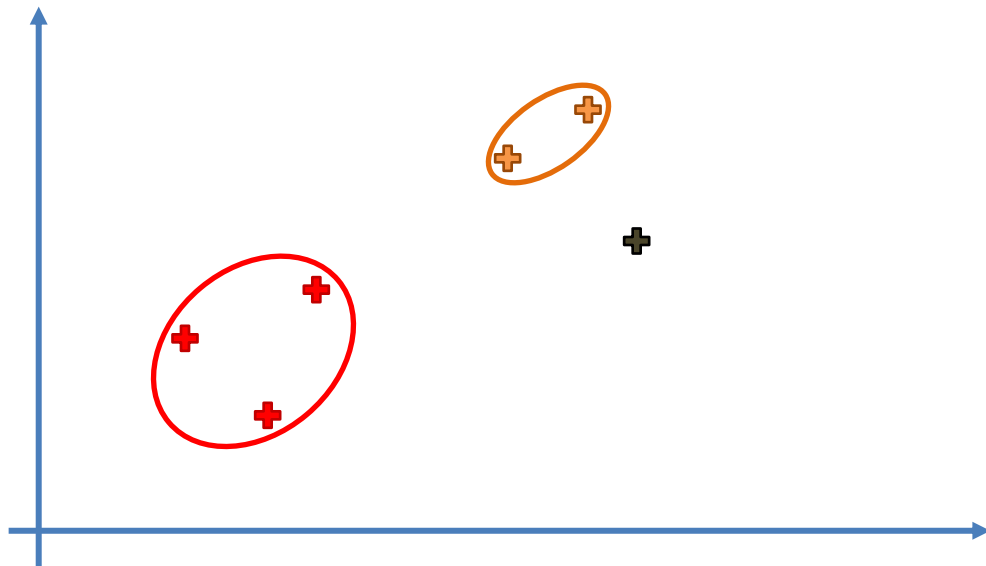
PASO 3: Elegir los dos clusters más cercanos y juntarlos en un único cluster

→ Así tenemos 4 clusters



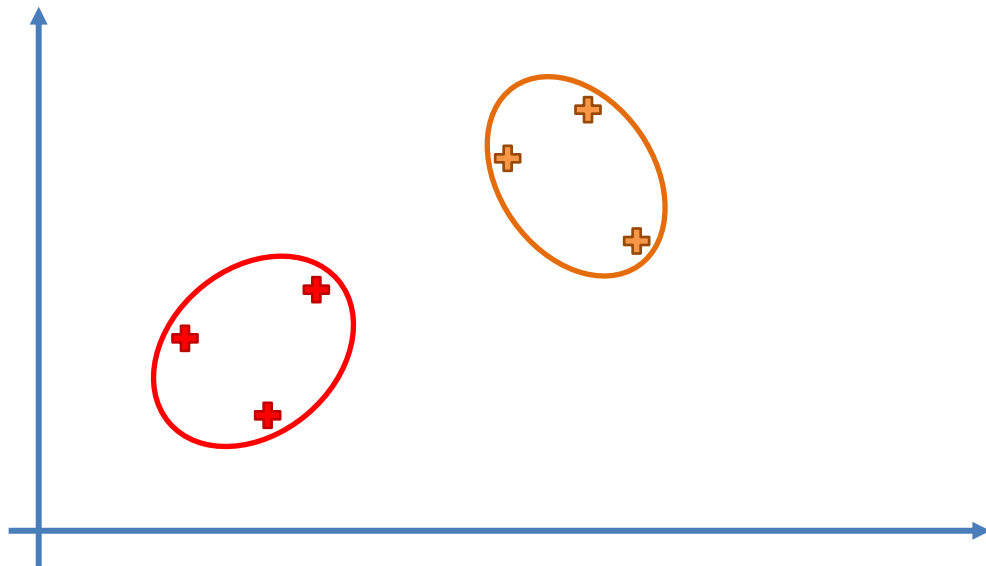
Clustering Jerárquico Aglomerativo

PASO 4: Repetir el PASO 3 hasta que quede un solo cluster



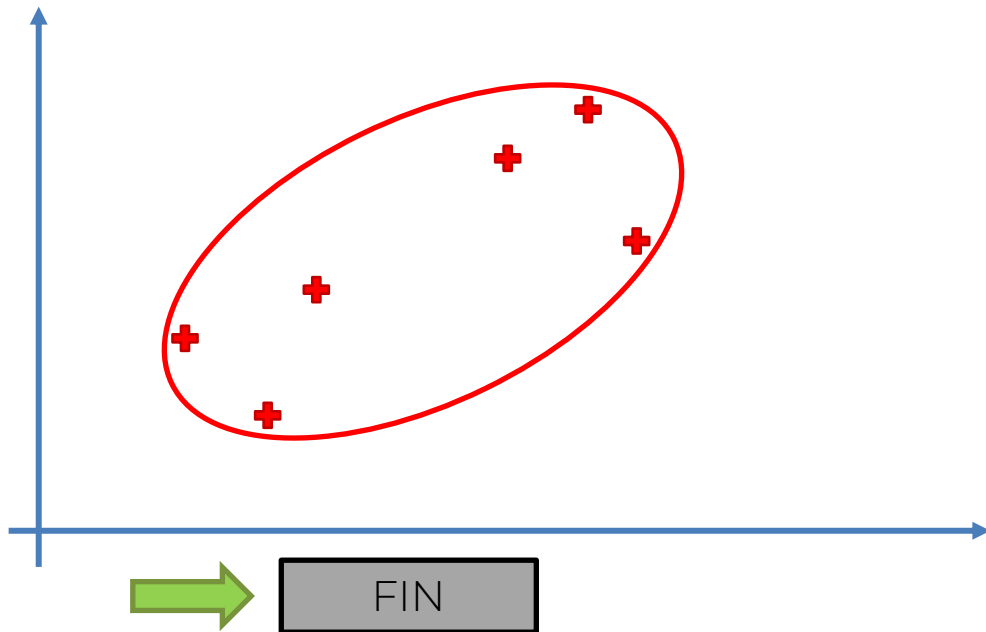
Clustering Jerárquico Aglomerativo

PASO 4: Repetir el PASO 3 hasta que quede un solo cluster



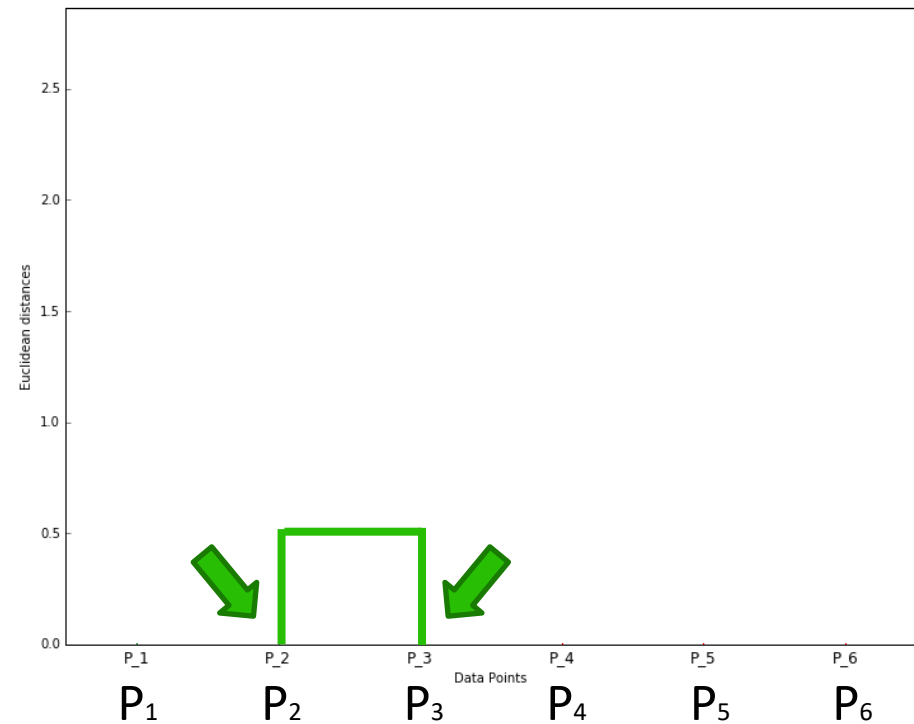
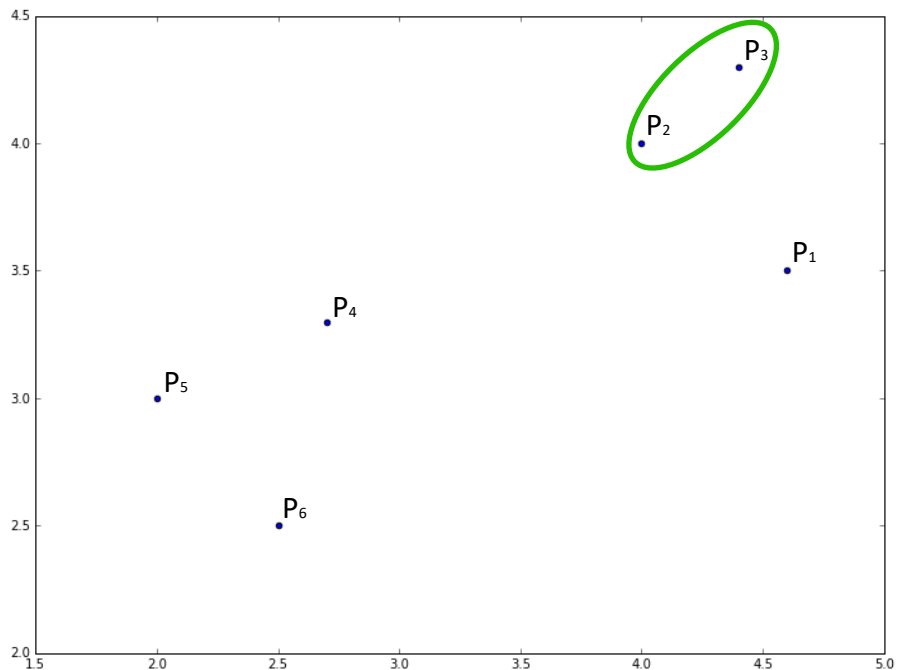
Clustering Jerárquico Aglomerativo

PASO 4: Repetir el PASO 3 hasta que quede un solo cluster

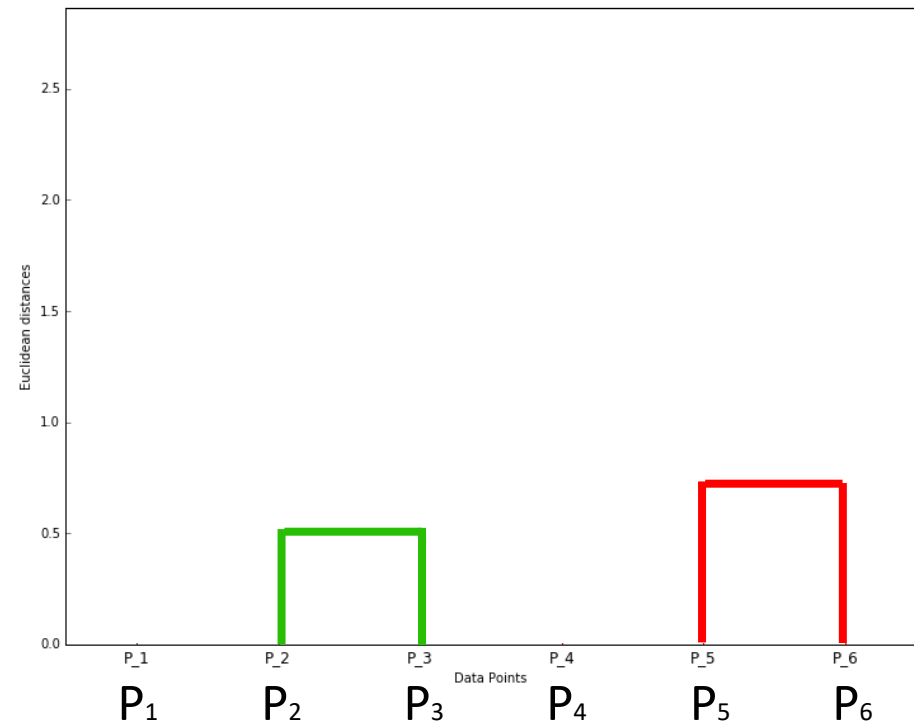
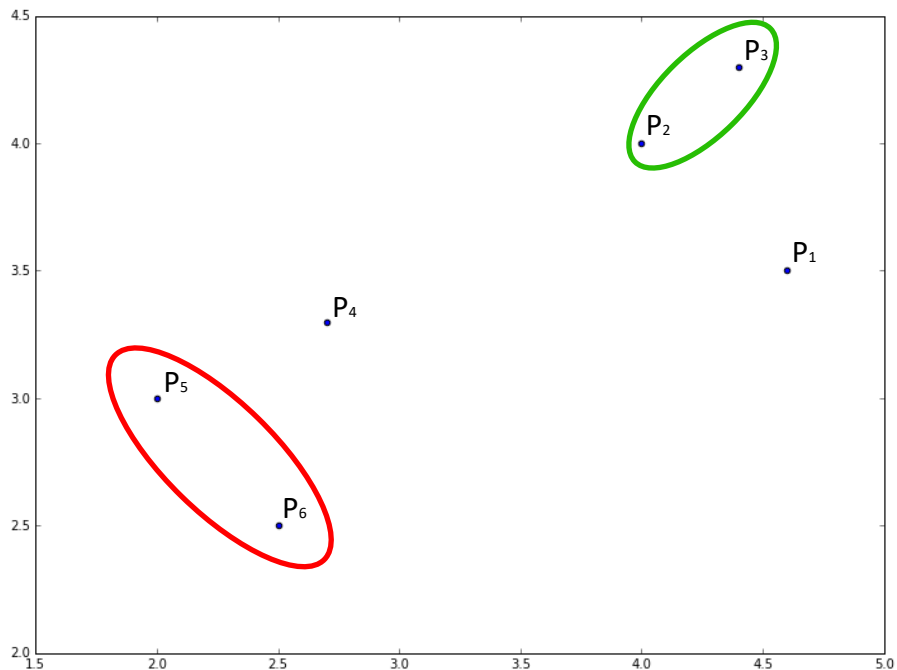


Idea del Clustering Jerárquico: ¿Cómo funcionan los dendrogramas?

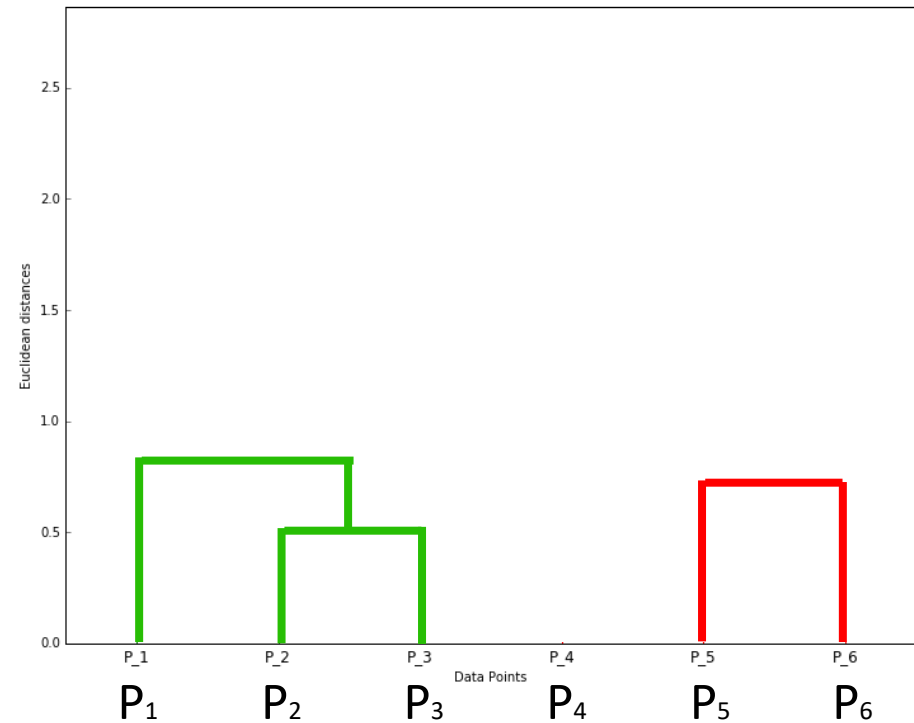
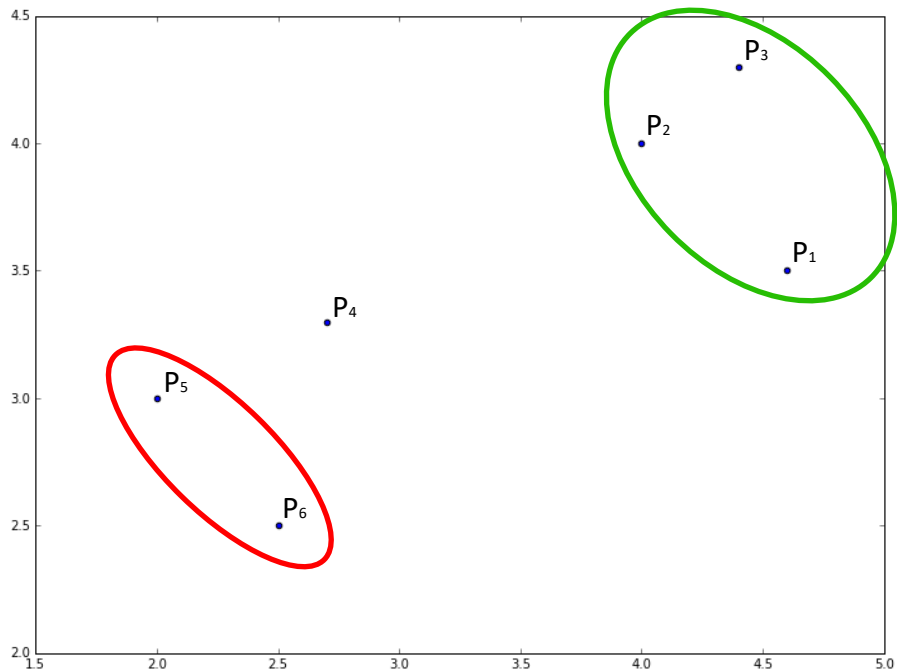
¿Cómo funcionan los dendrogramas?



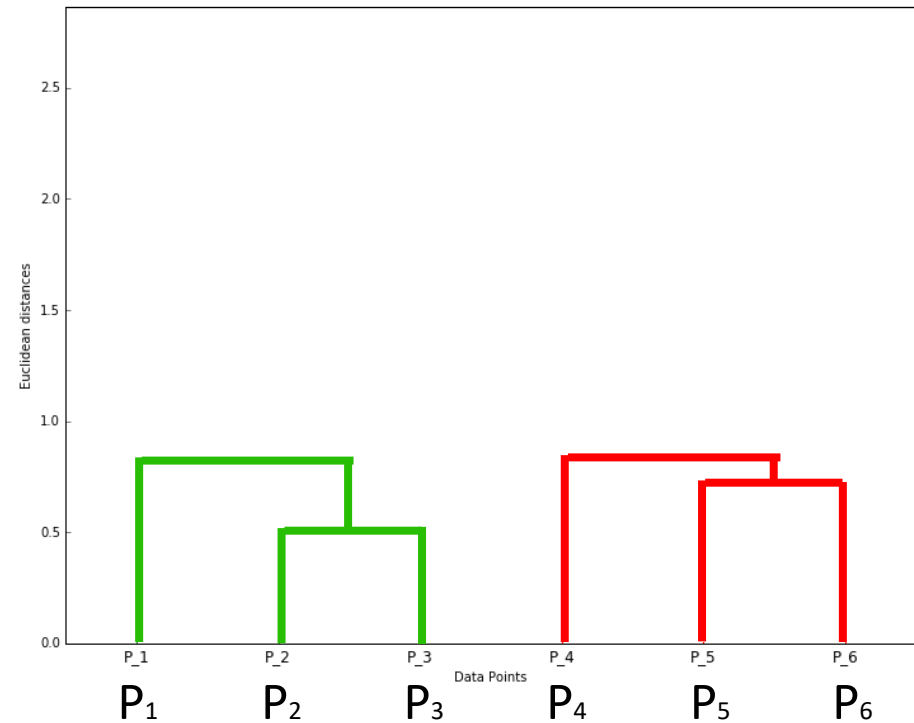
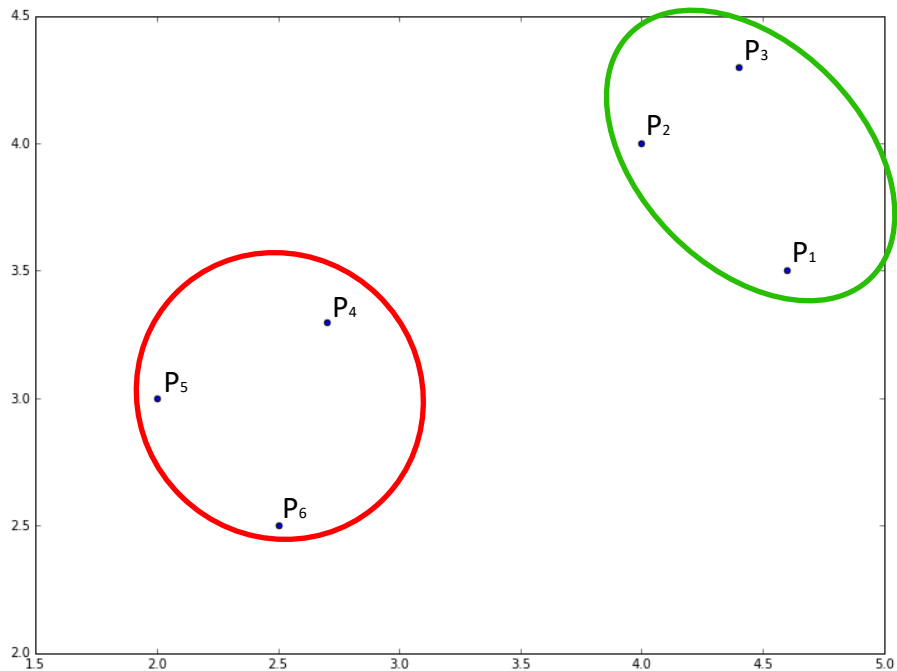
¿Cómo funcionan los dendrogramas?



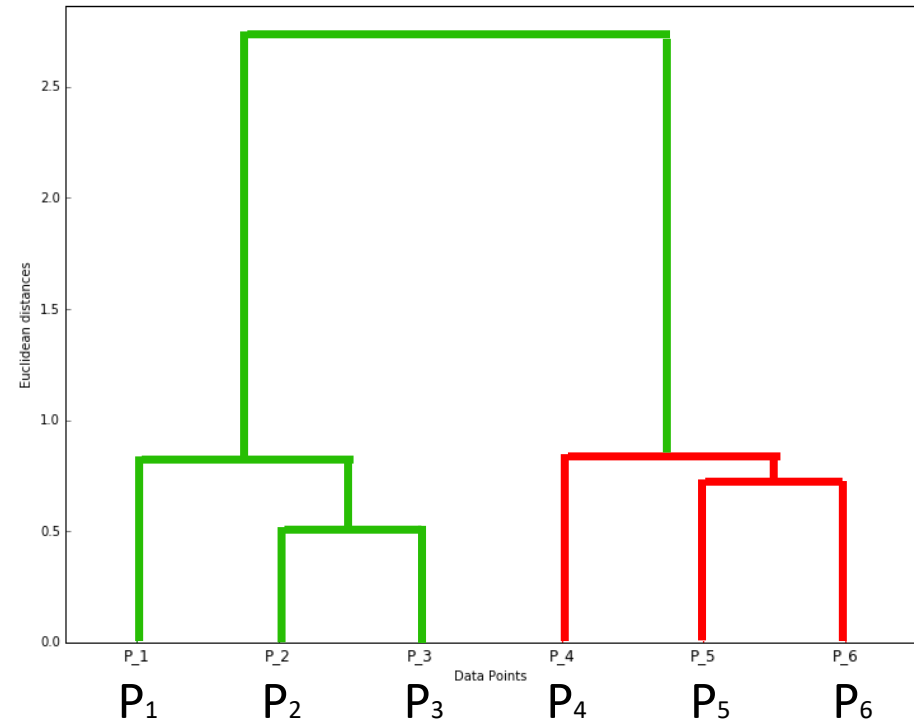
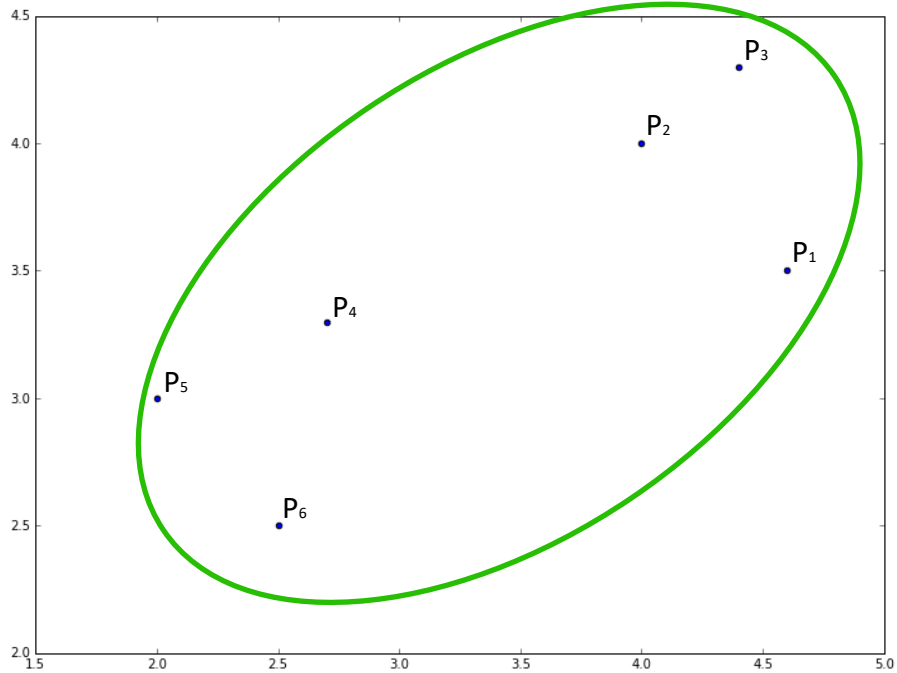
¿Cómo funcionan los dendrogramas?



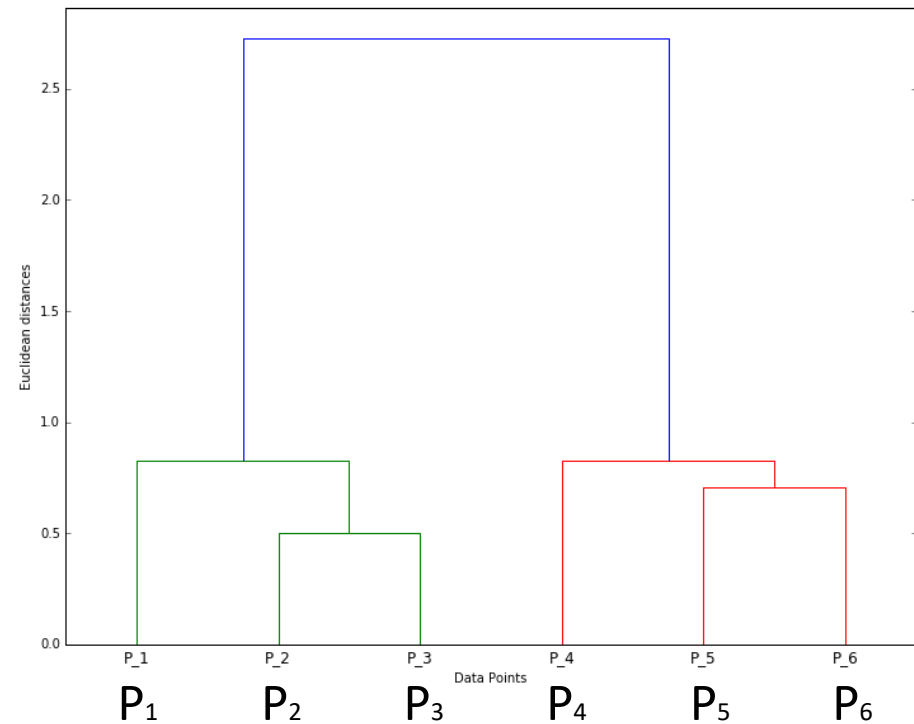
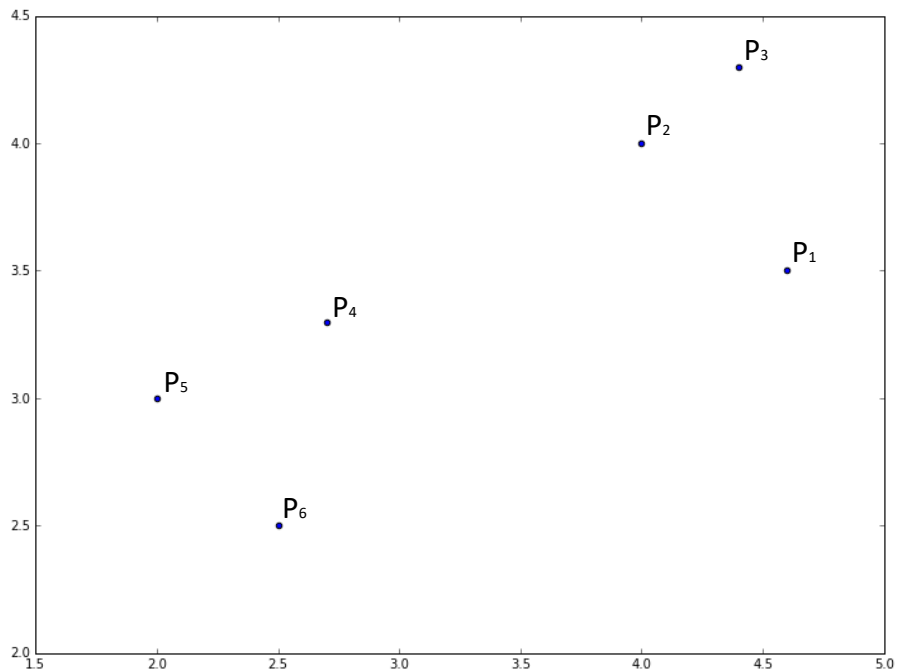
¿Cómo funcionan los dendrogramas?



¿Cómo funcionan los dendrogramas?



¿Cómo funcionan los dendrogramas?



VAMOS
AL
CÓDIGO!!