CS 156:Introduction to Artificial Intelligence

Instructor: Dr. Sayma Akther

San José State University

Clustering Algorithms Type: K-Means Algorithm

01

K-Means algorithm is one of the most commonly used clustering techniques. It divides the samples into separate clusters with equal variances to classify the data

02

This approach requires the number of clusters to be provided. It is fast with fewer computations needed, with the linear complexity of O(n)

K means clustering algorithm in unsupervised learning

K-Means clustering groups similar data points together into clusters, aims to find natural patterns in data by partitioning it into K distinct groups, until clusters stabilize

How does it work

Initiation

- Choose clusters (K) for creating data
- Deploy K centroids in dataset
- Add text here

Assignment

- Assign each data point to nearest cluster based on similarities (Euclidean distance)
- Add text here

Update centroids

- Recalculate cluster centroids by taking mean of data points assignments
- Add text here

Iterate

- Repeat until reaches convergence
- Occurs when assignments and centroids no longer change/are unstable
- Add text here

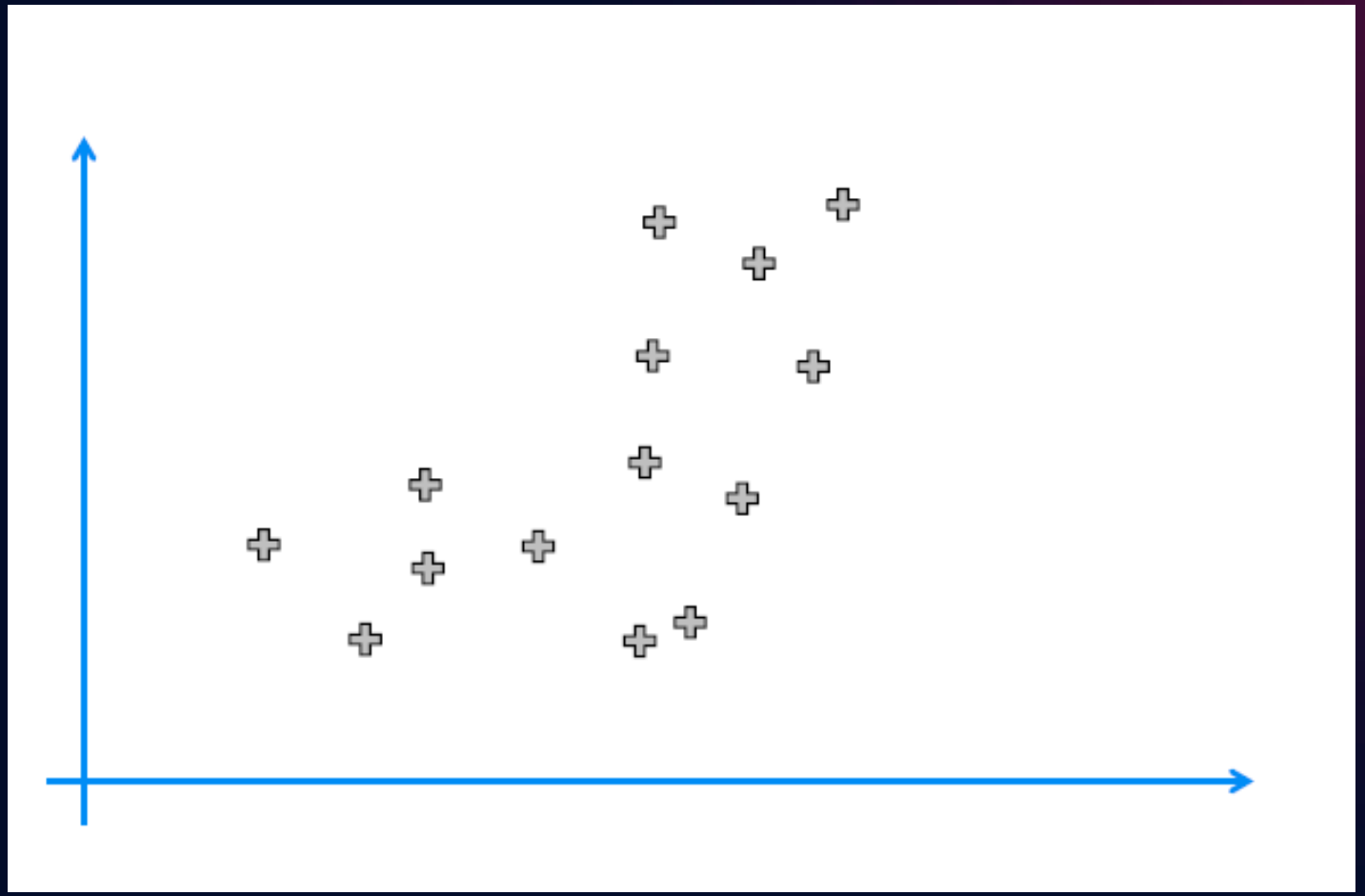
Major use cases

Academic performance tracking for smart results prediction and classroom management

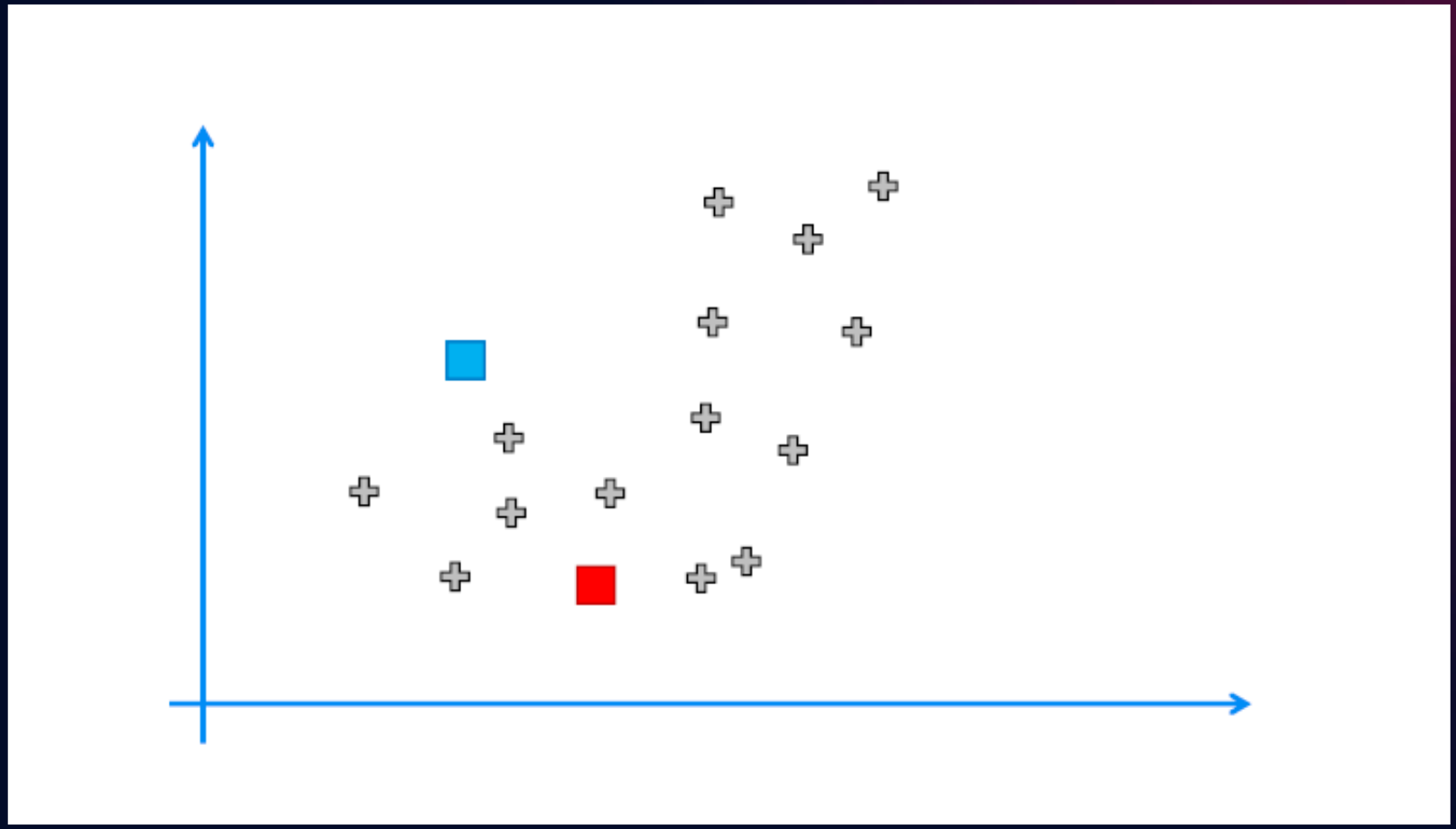
Search engines data clustering for most accurate results

Document clustering for compressed and smart information management

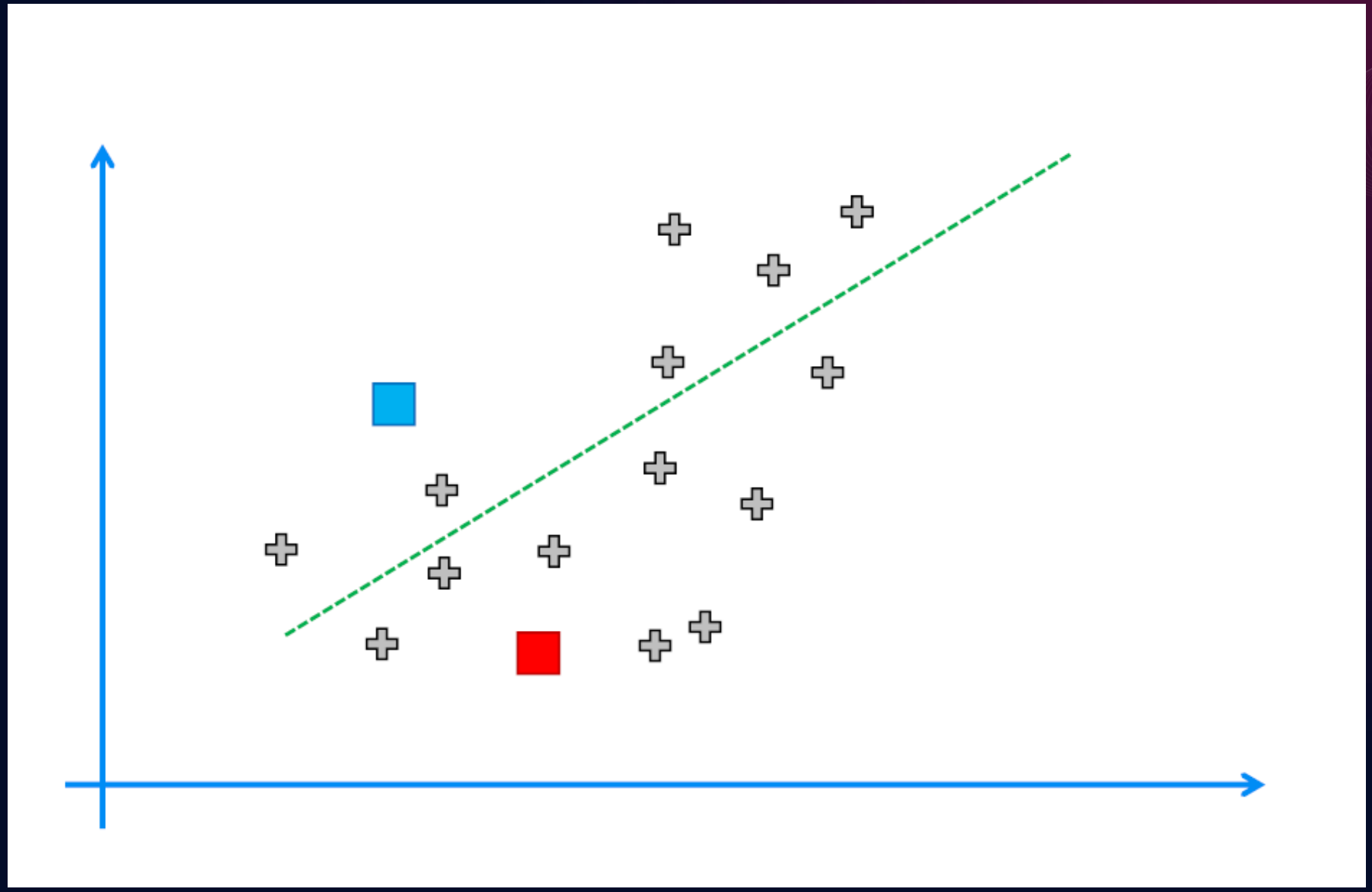
K means clustering algorithm



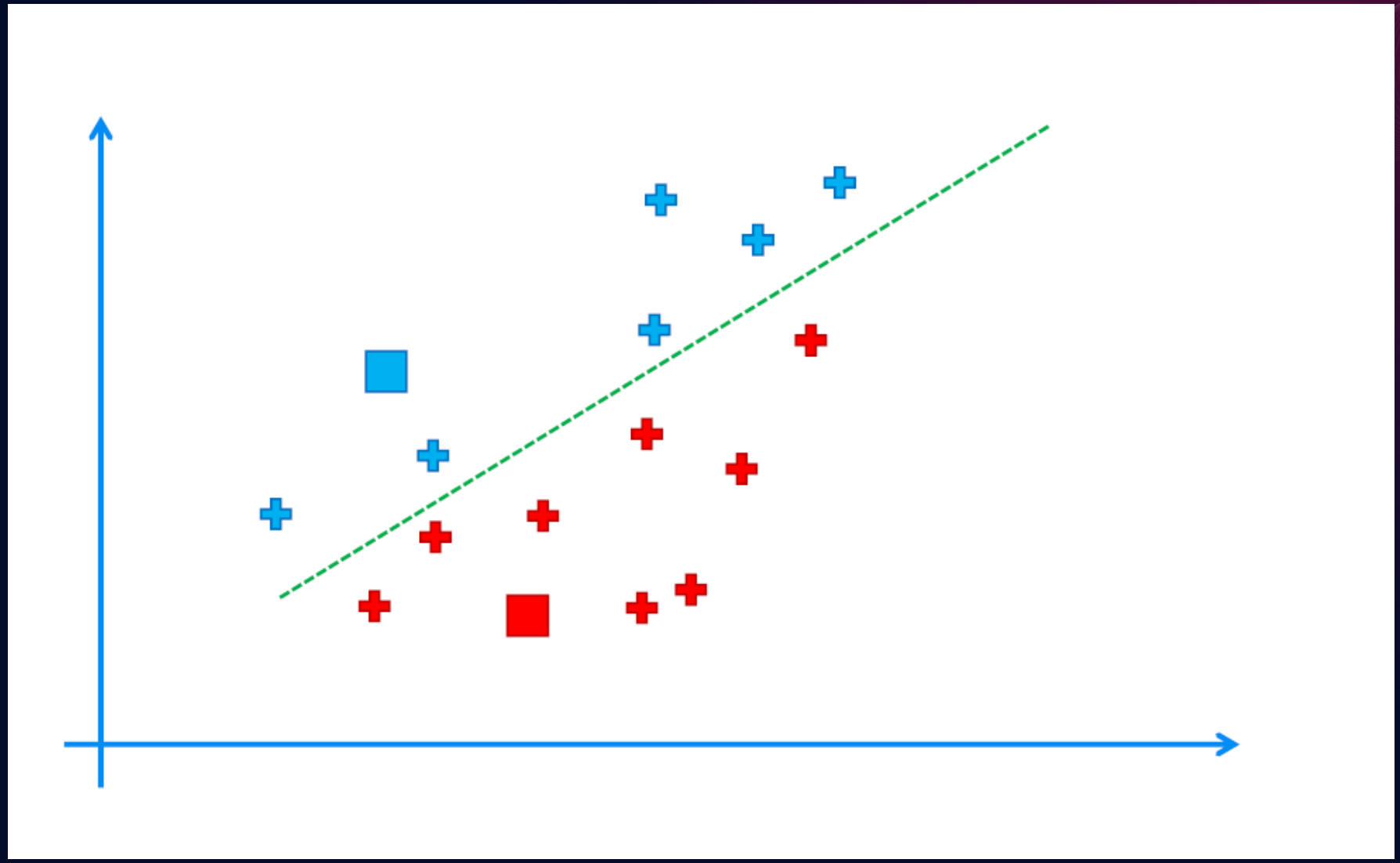
K means clustering algorithm



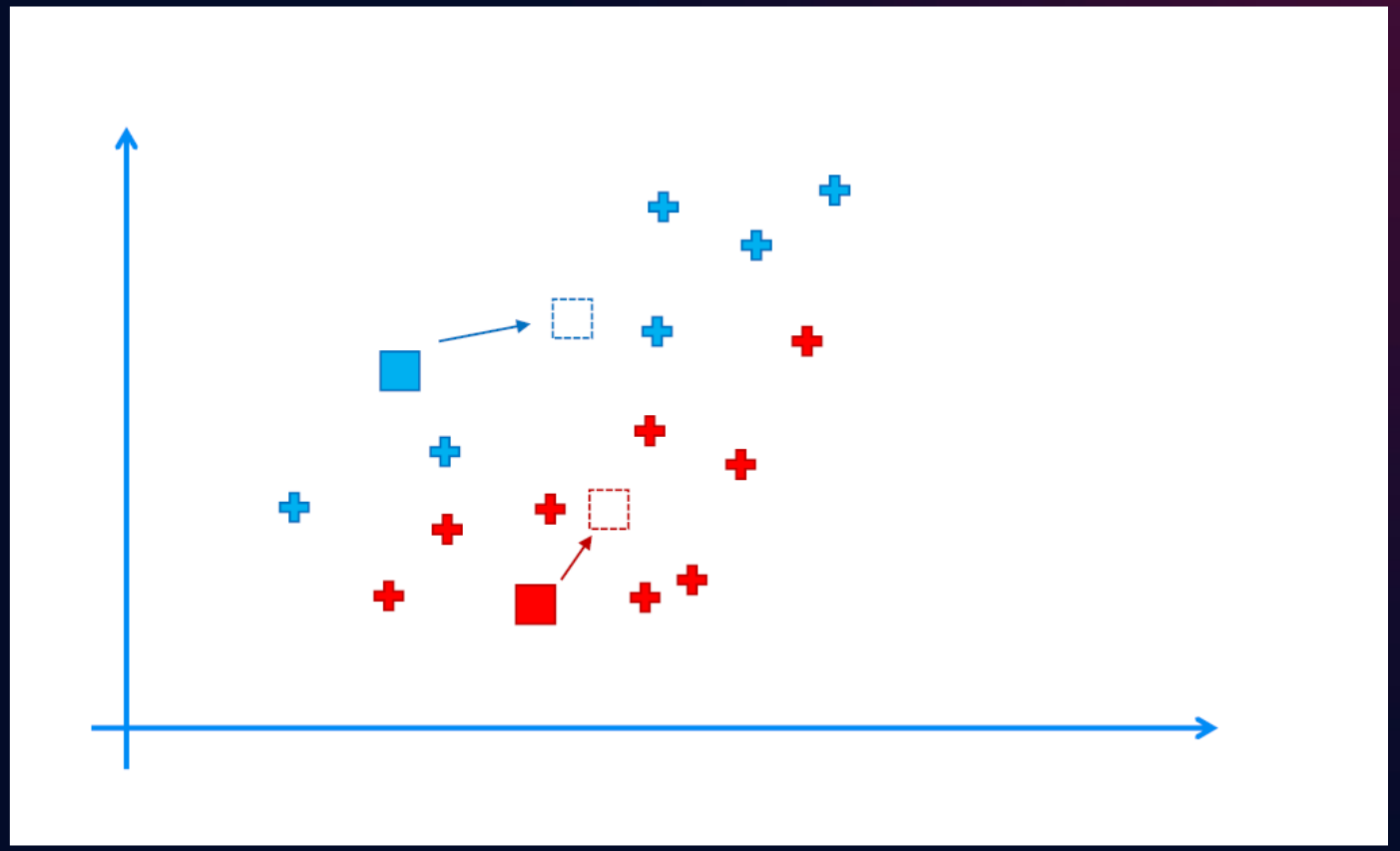
K means clustering algorithm



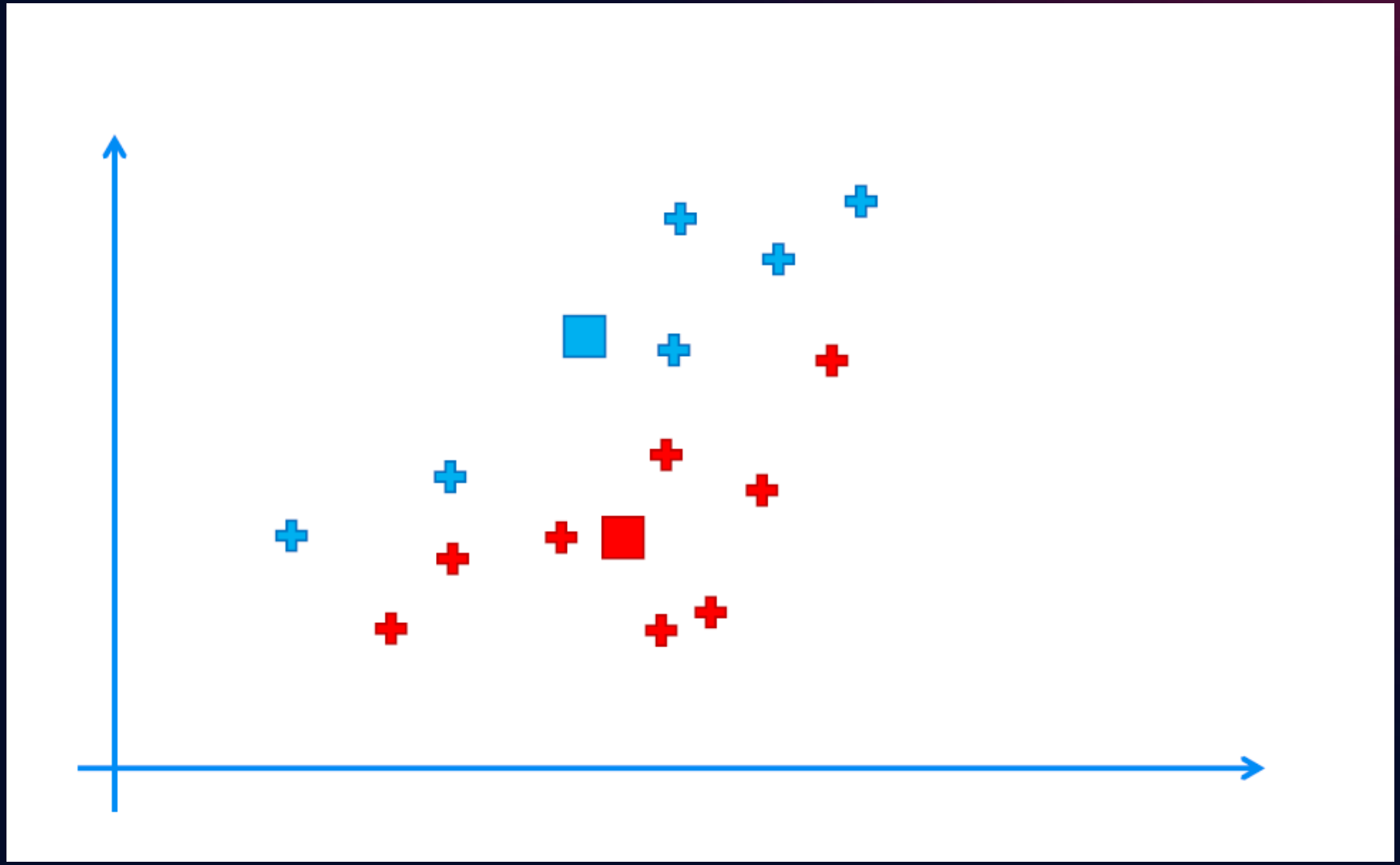
K means clustering algorithm



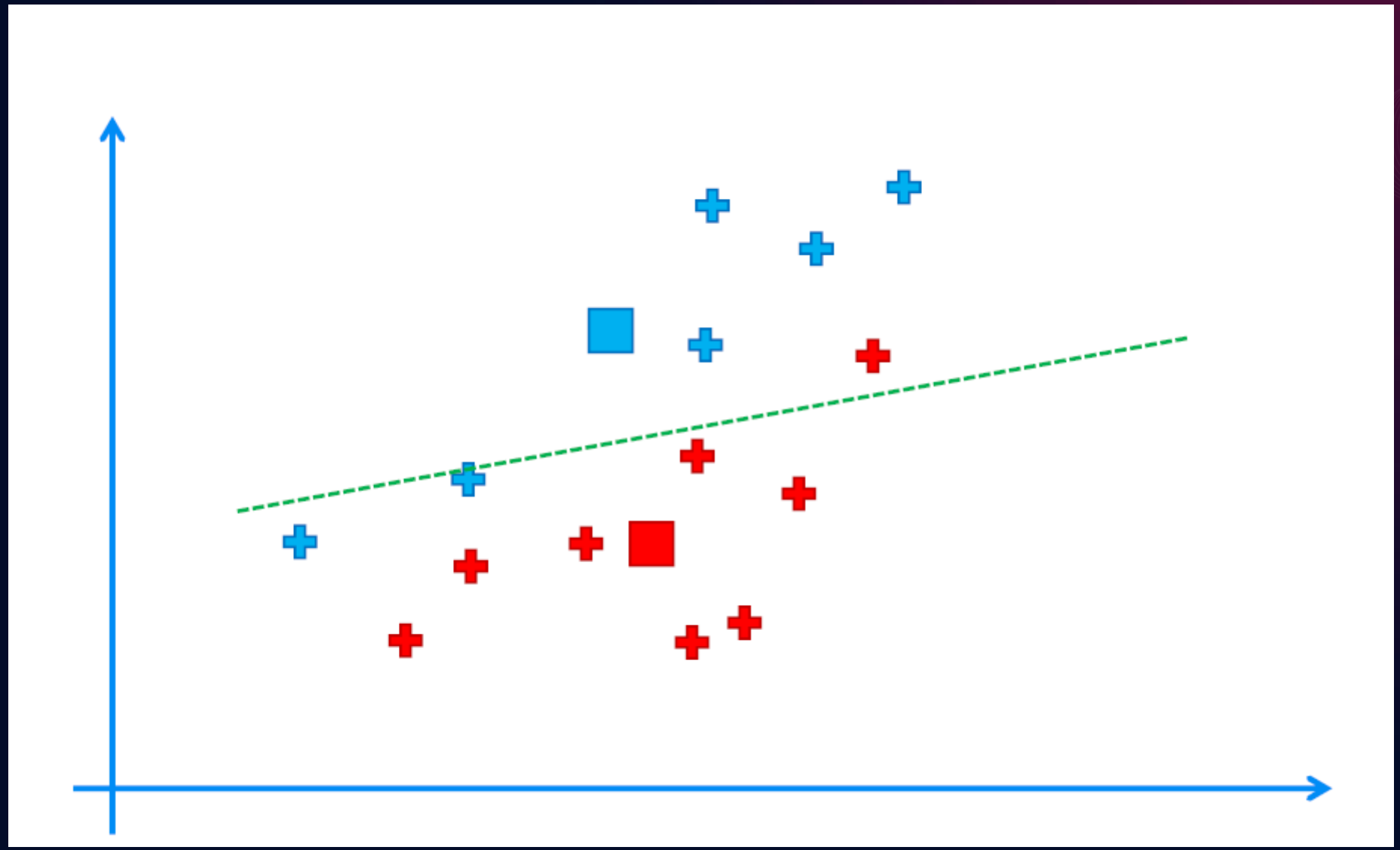
K means clustering algorithm



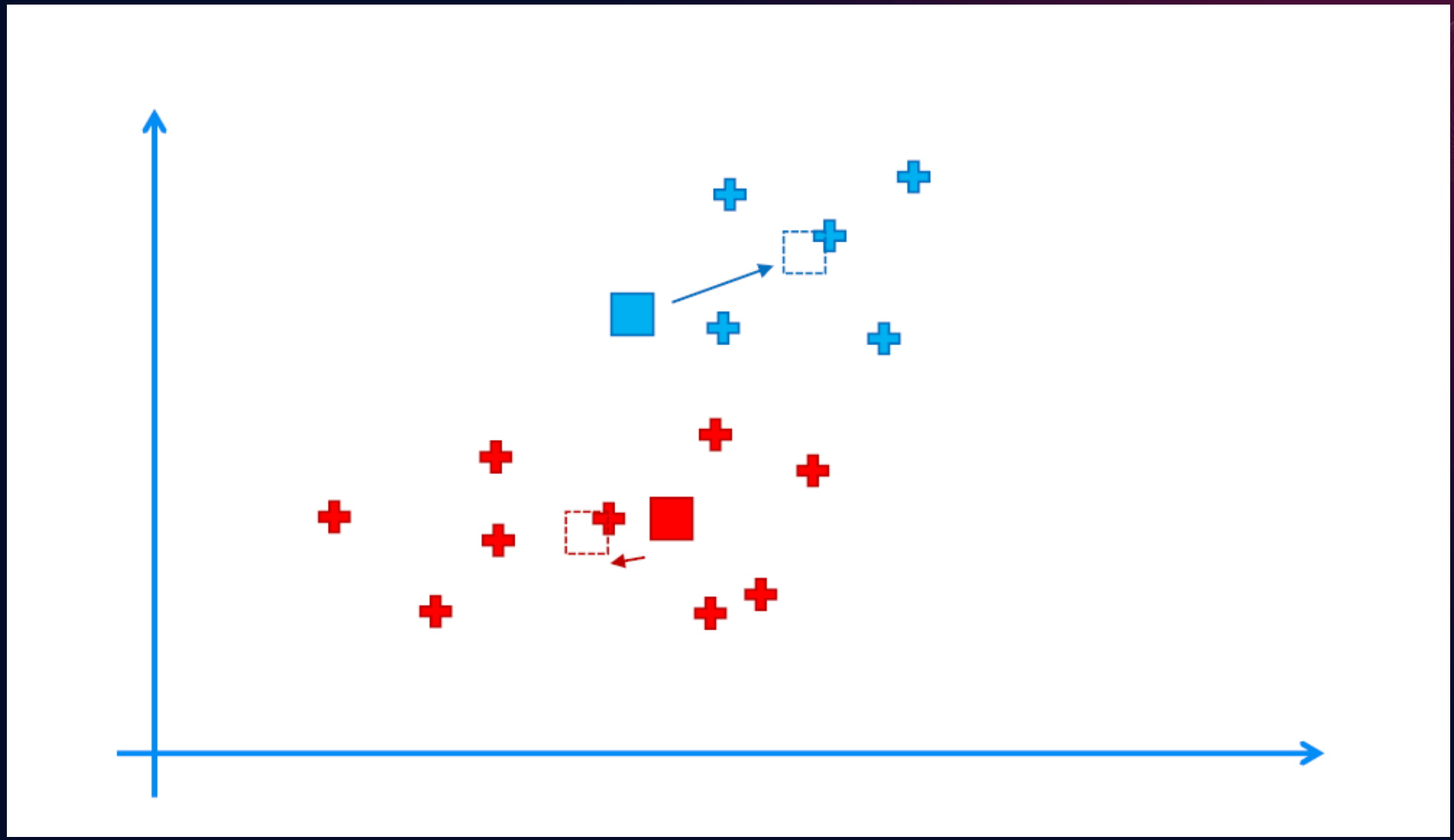
K means clustering algorithm



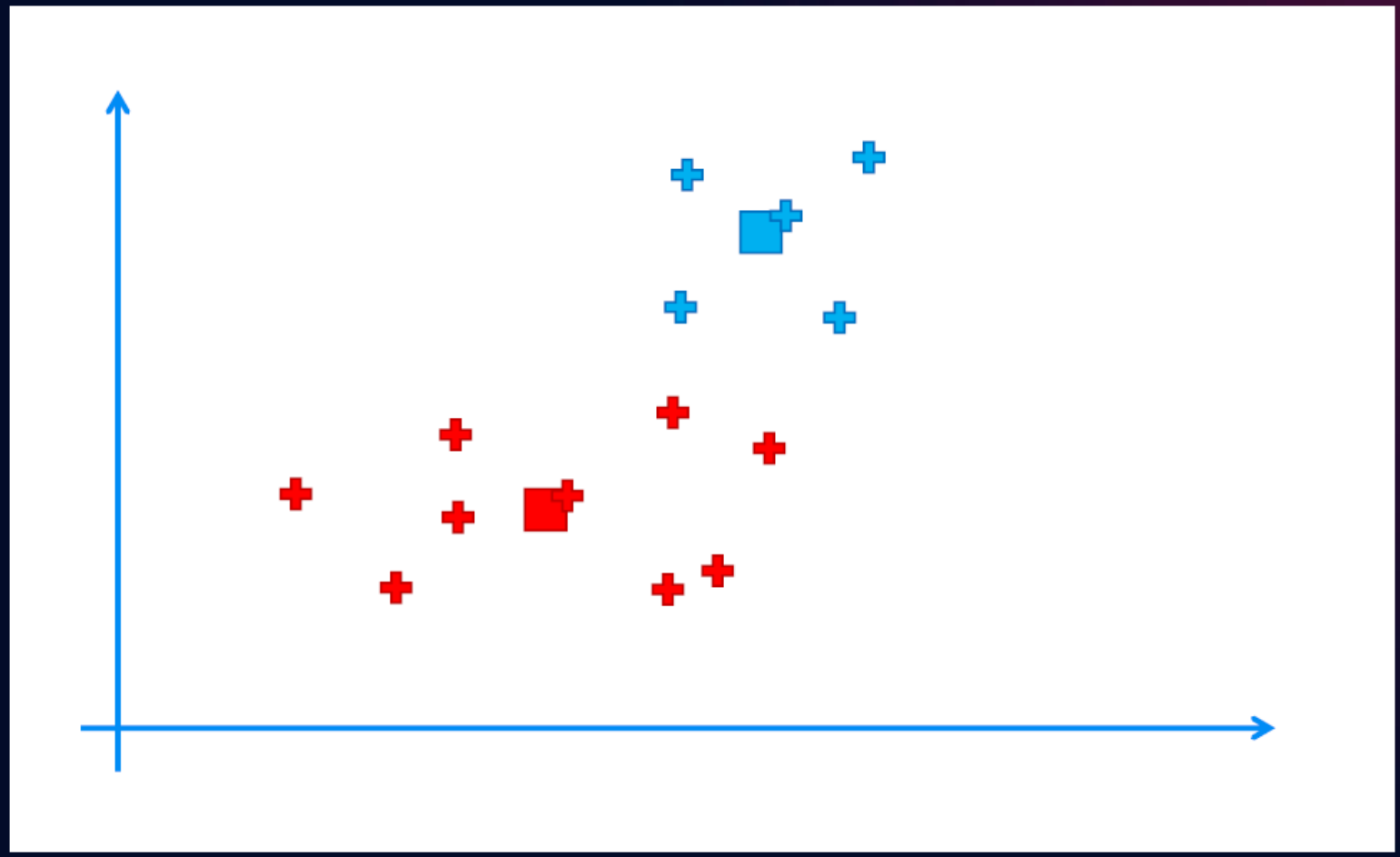
K means clustering algorithm



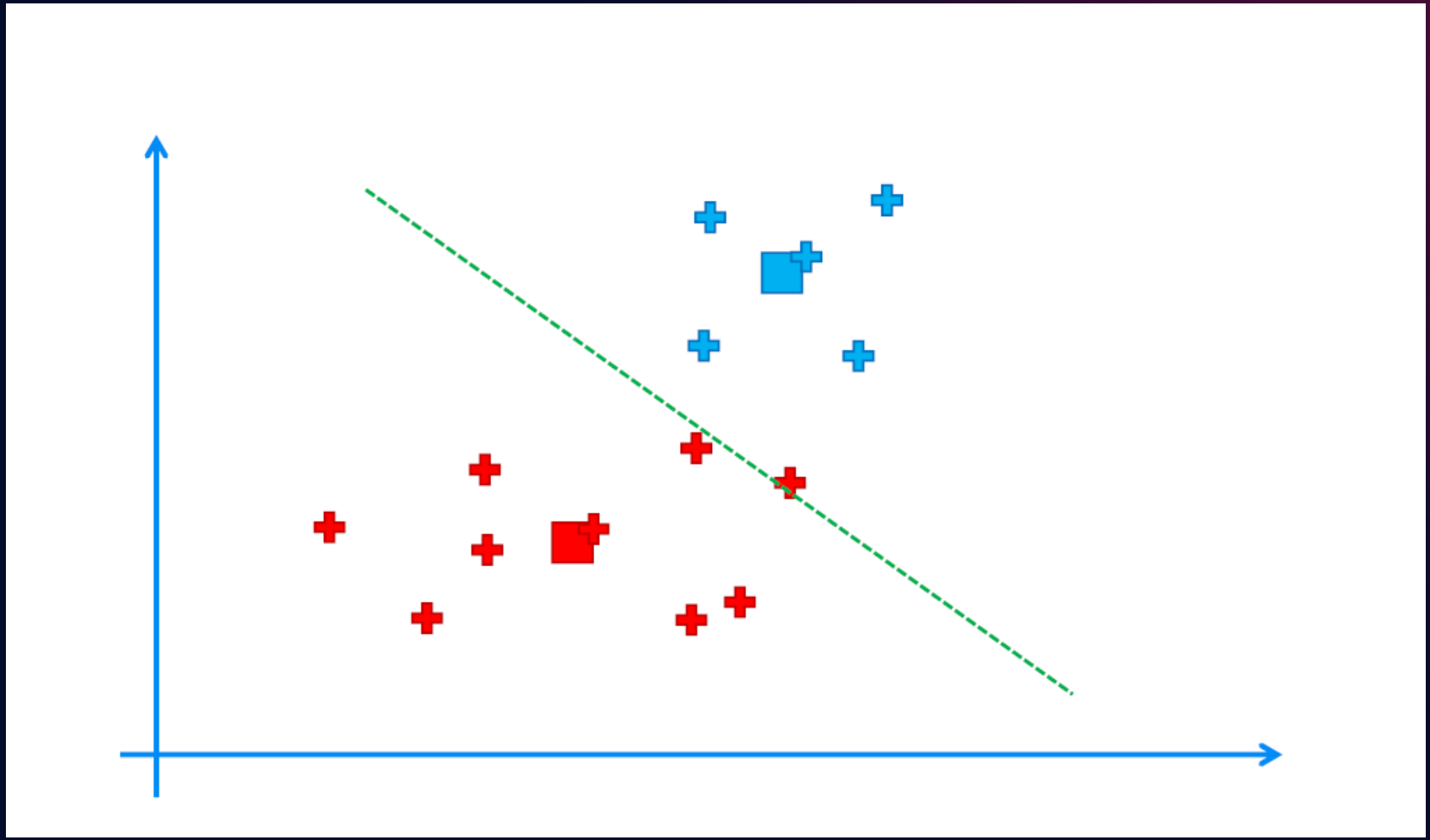
K means clustering algorithm



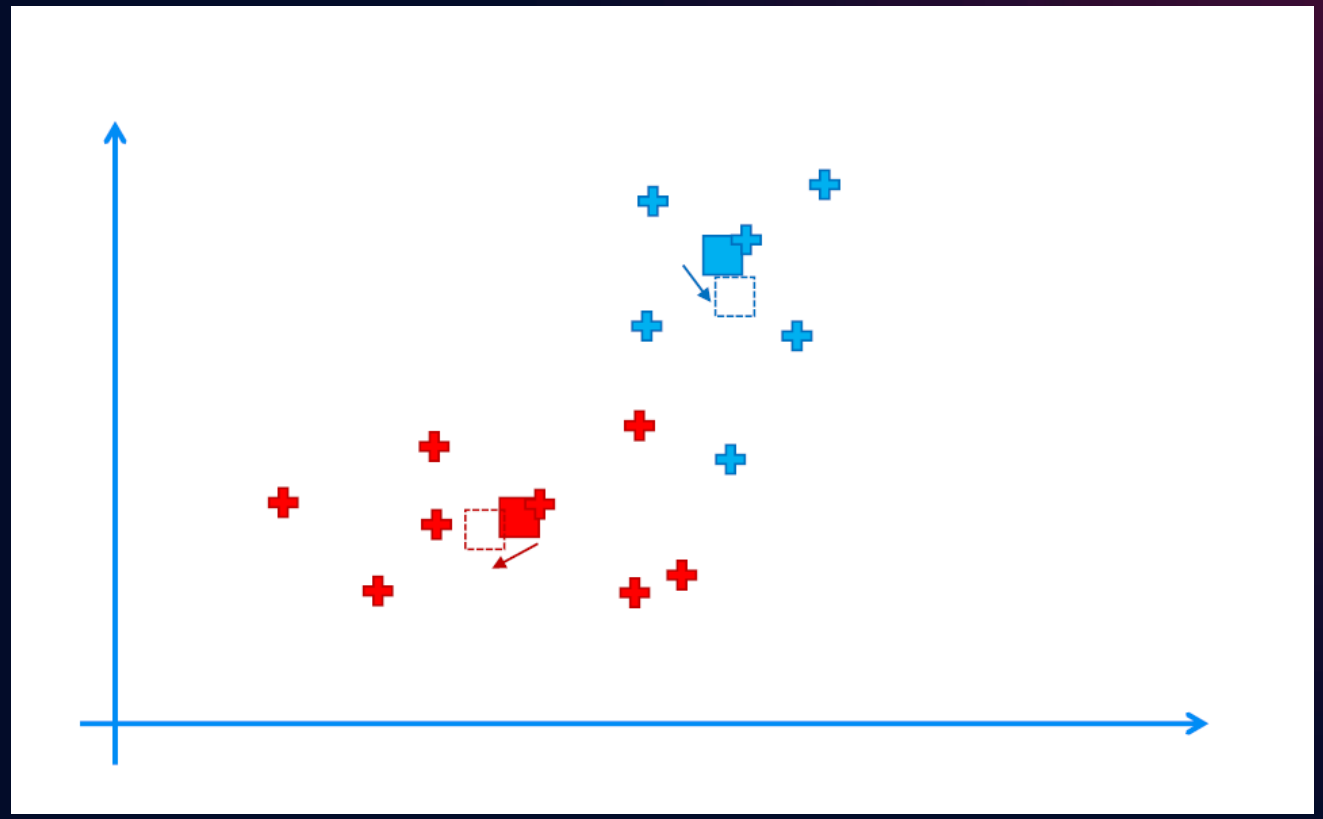
K means clustering algorithm



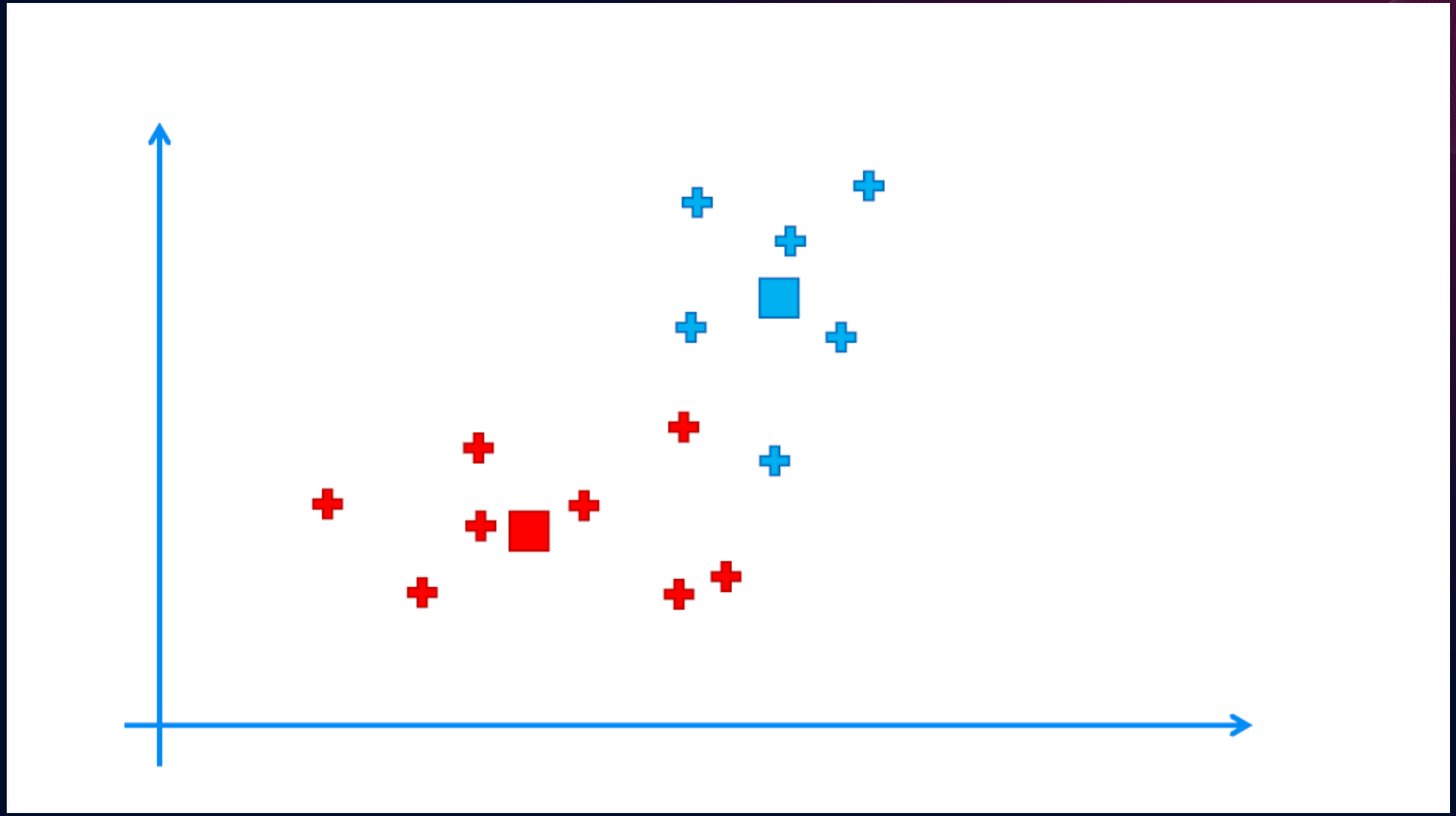
K means clustering algorithm



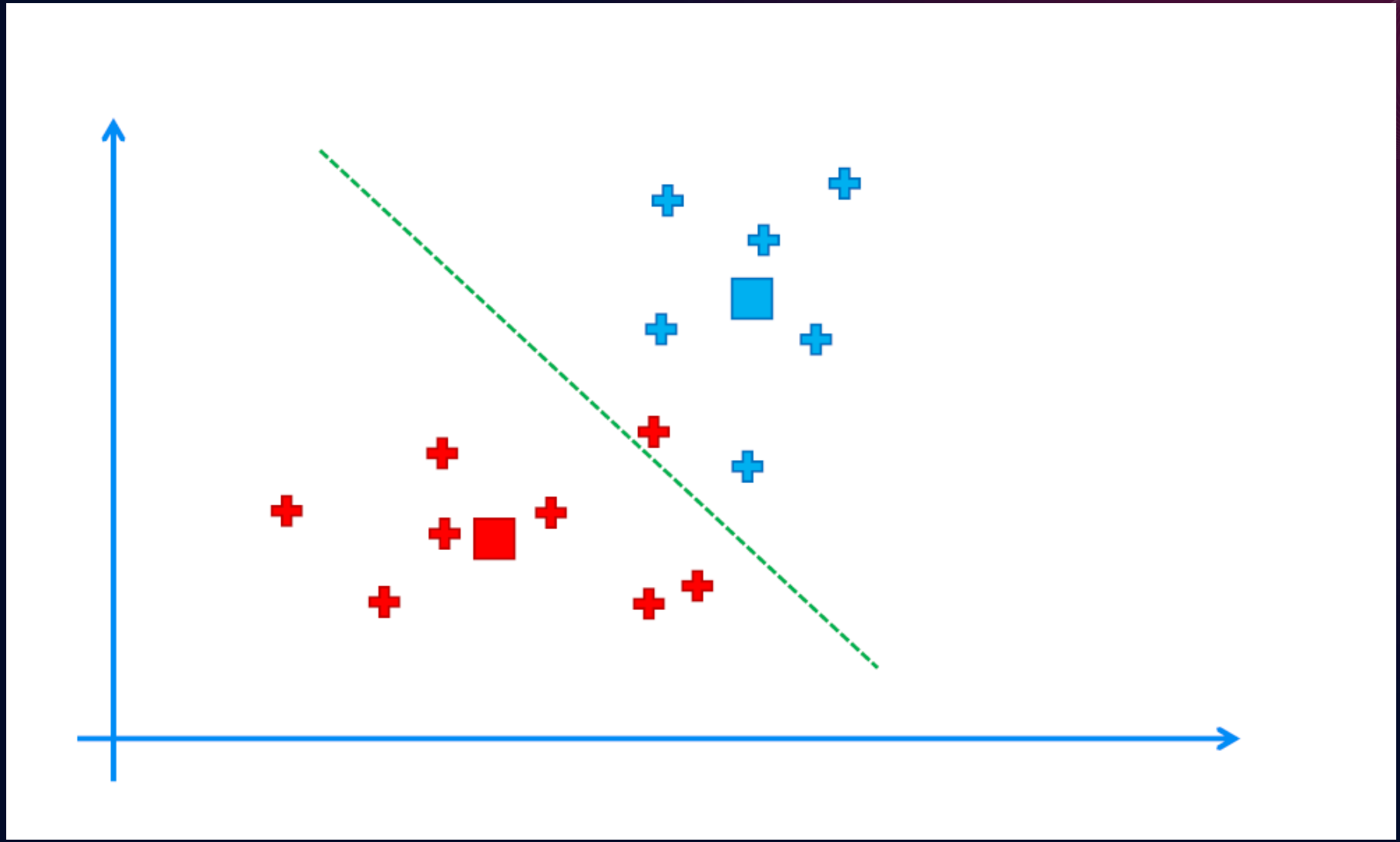
K means clustering algorithm



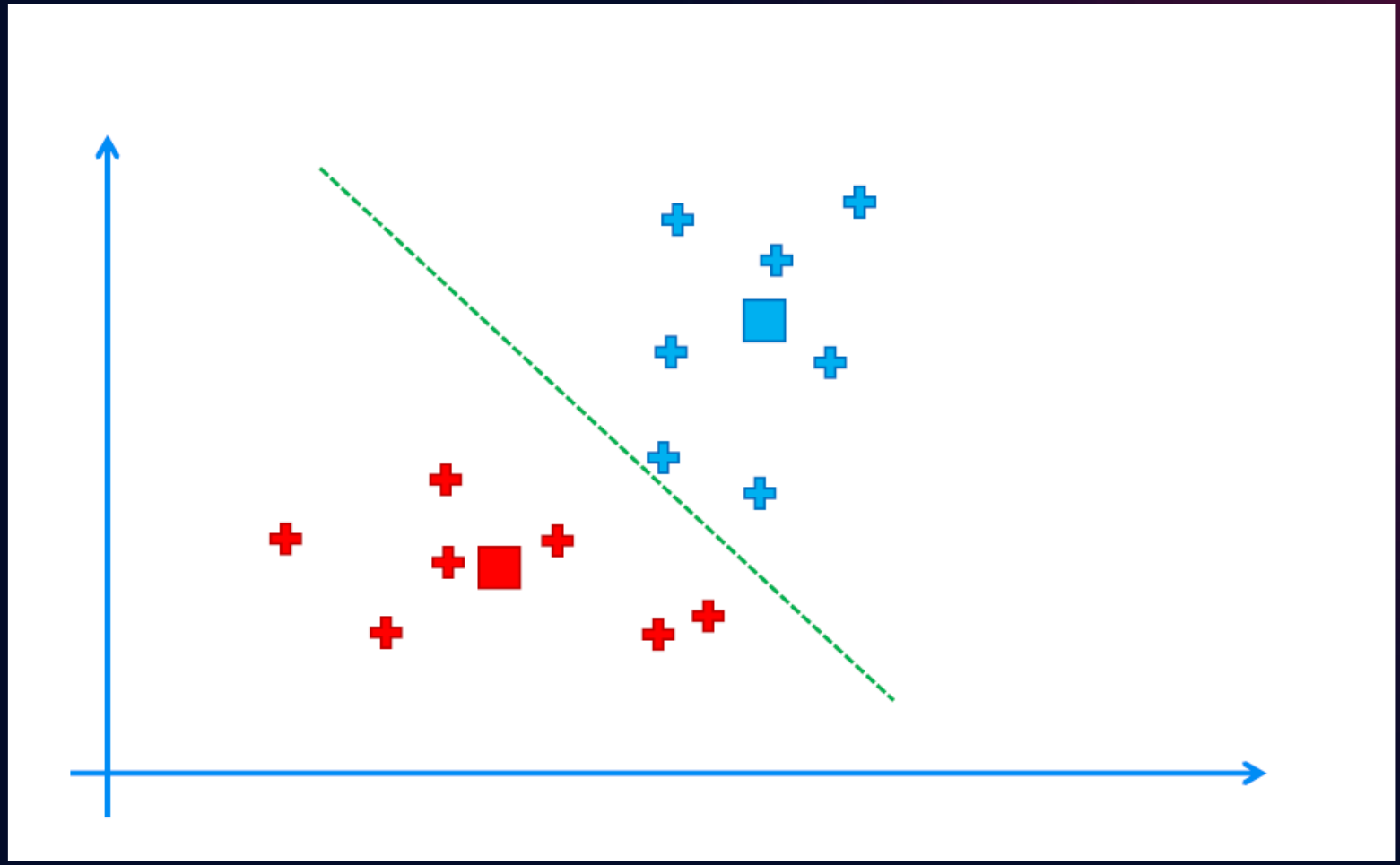
K means clustering algorithm



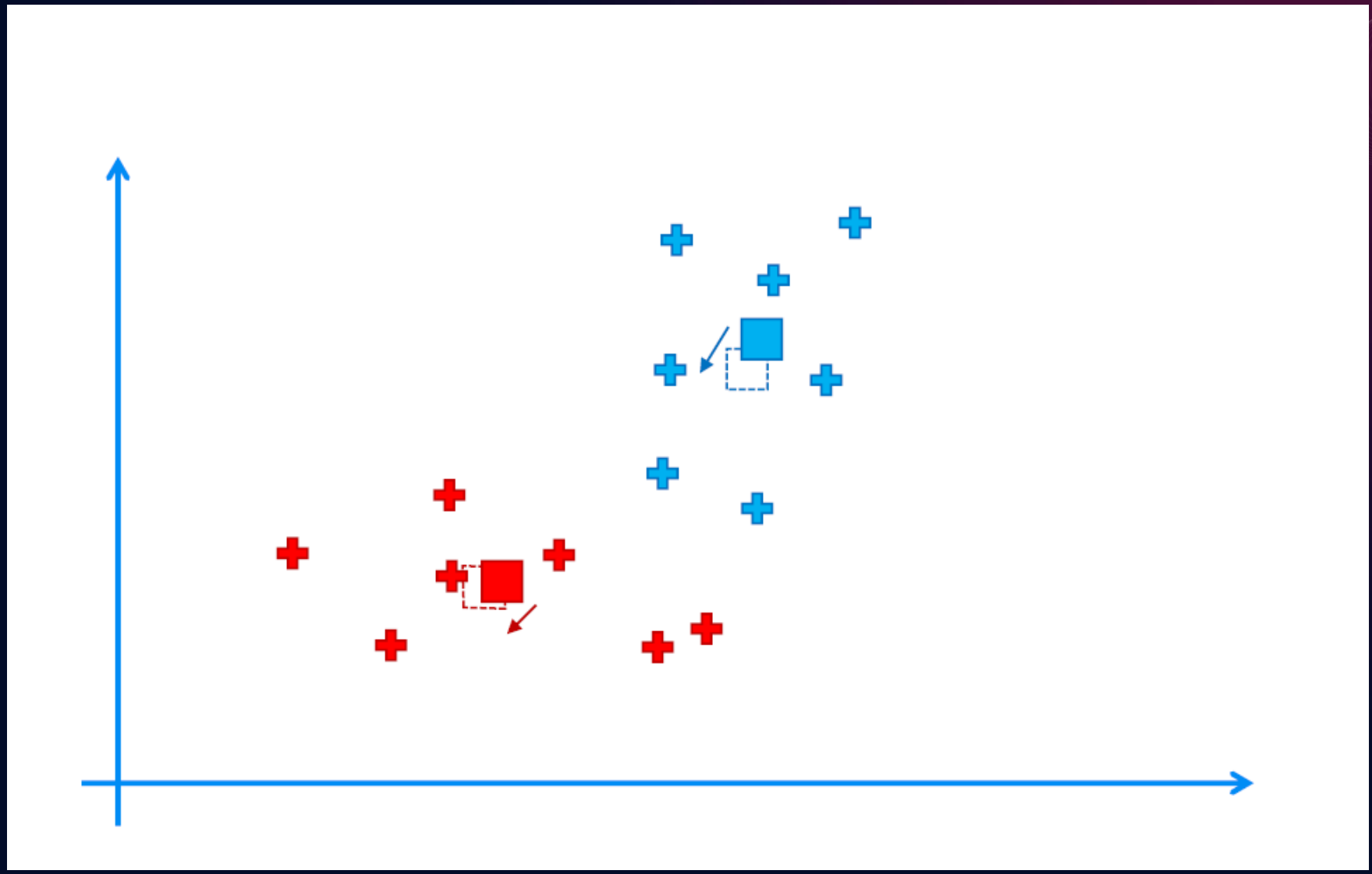
K means clustering algorithm



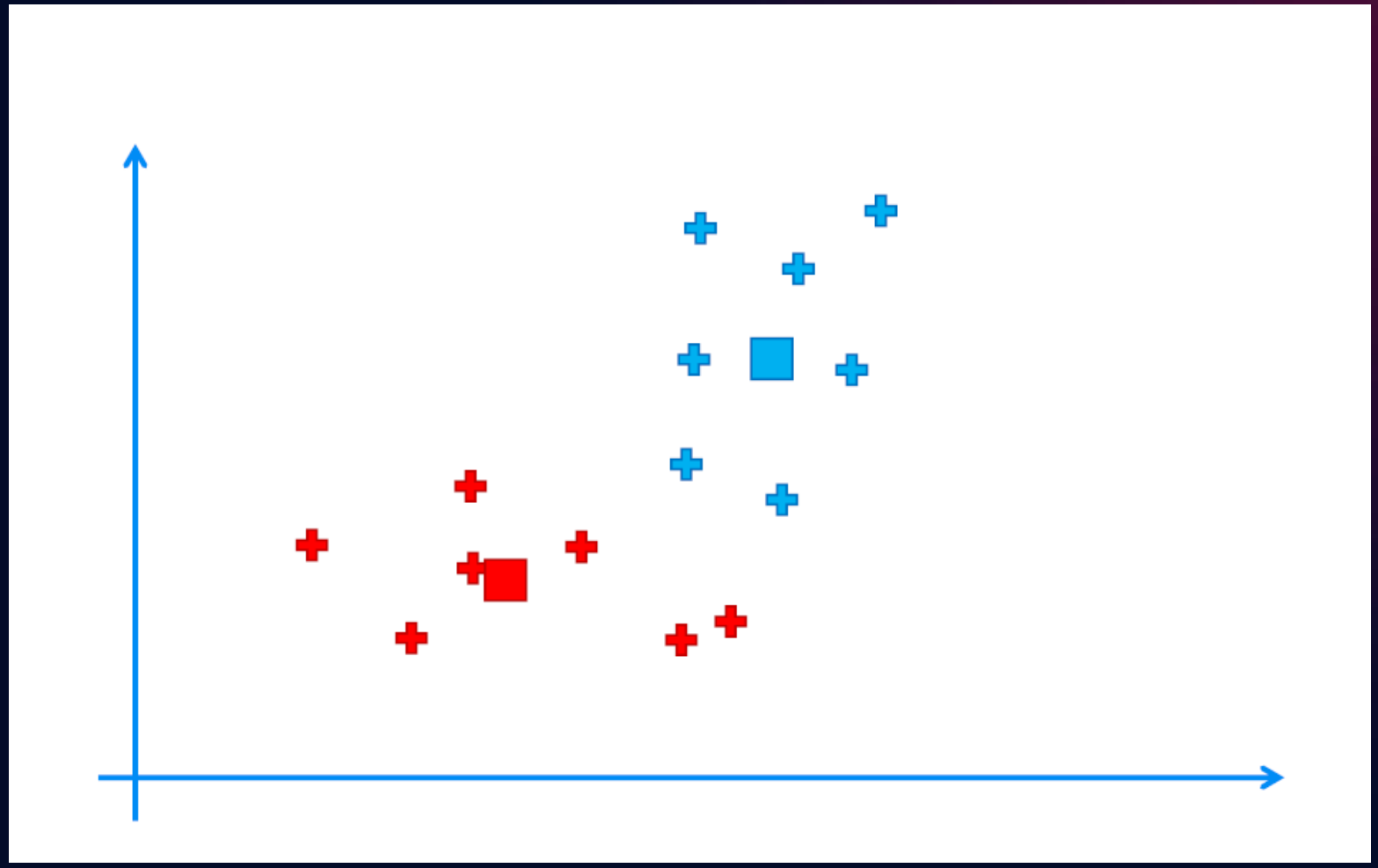
K means clustering algorithm



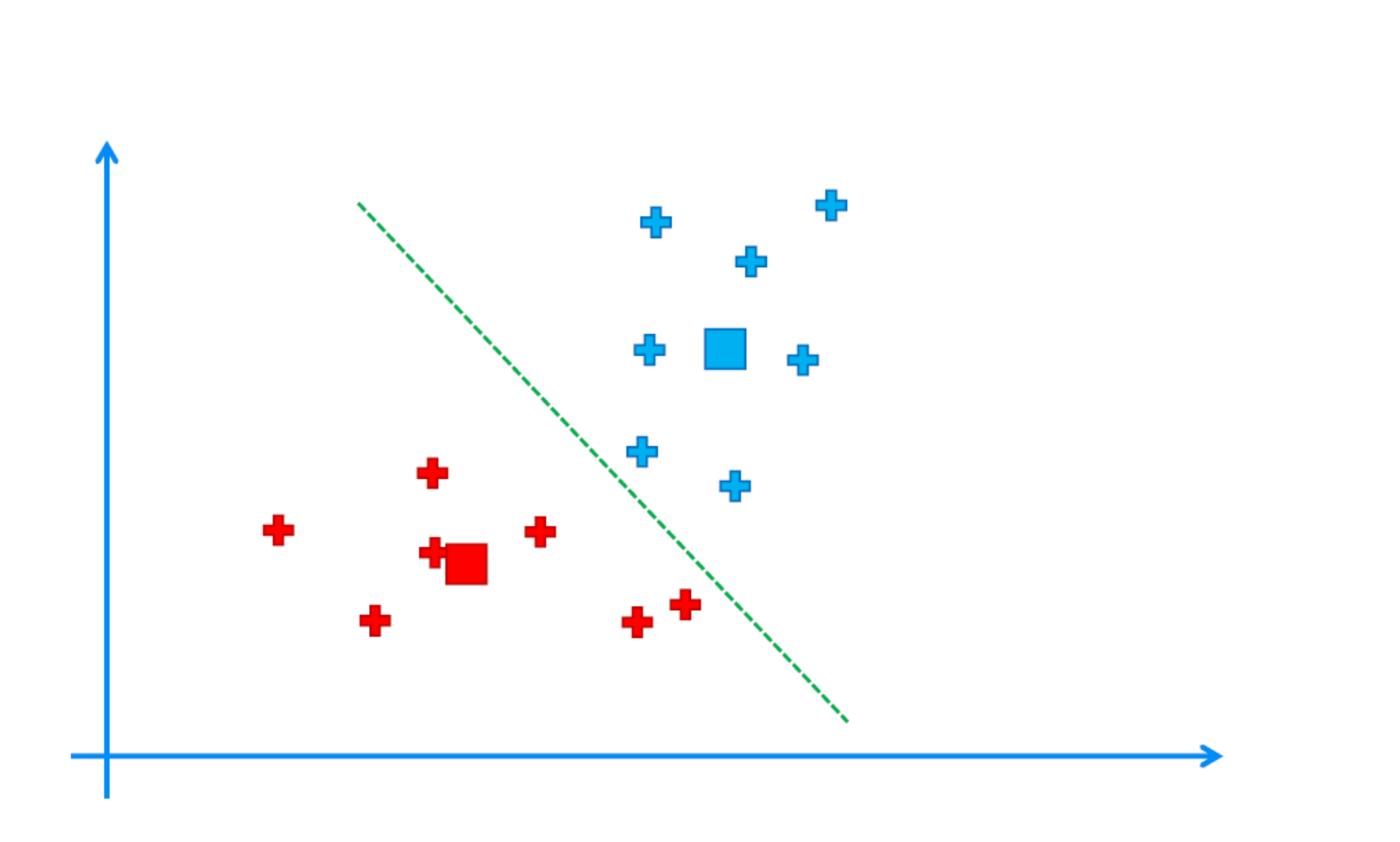
K means clustering algorithm



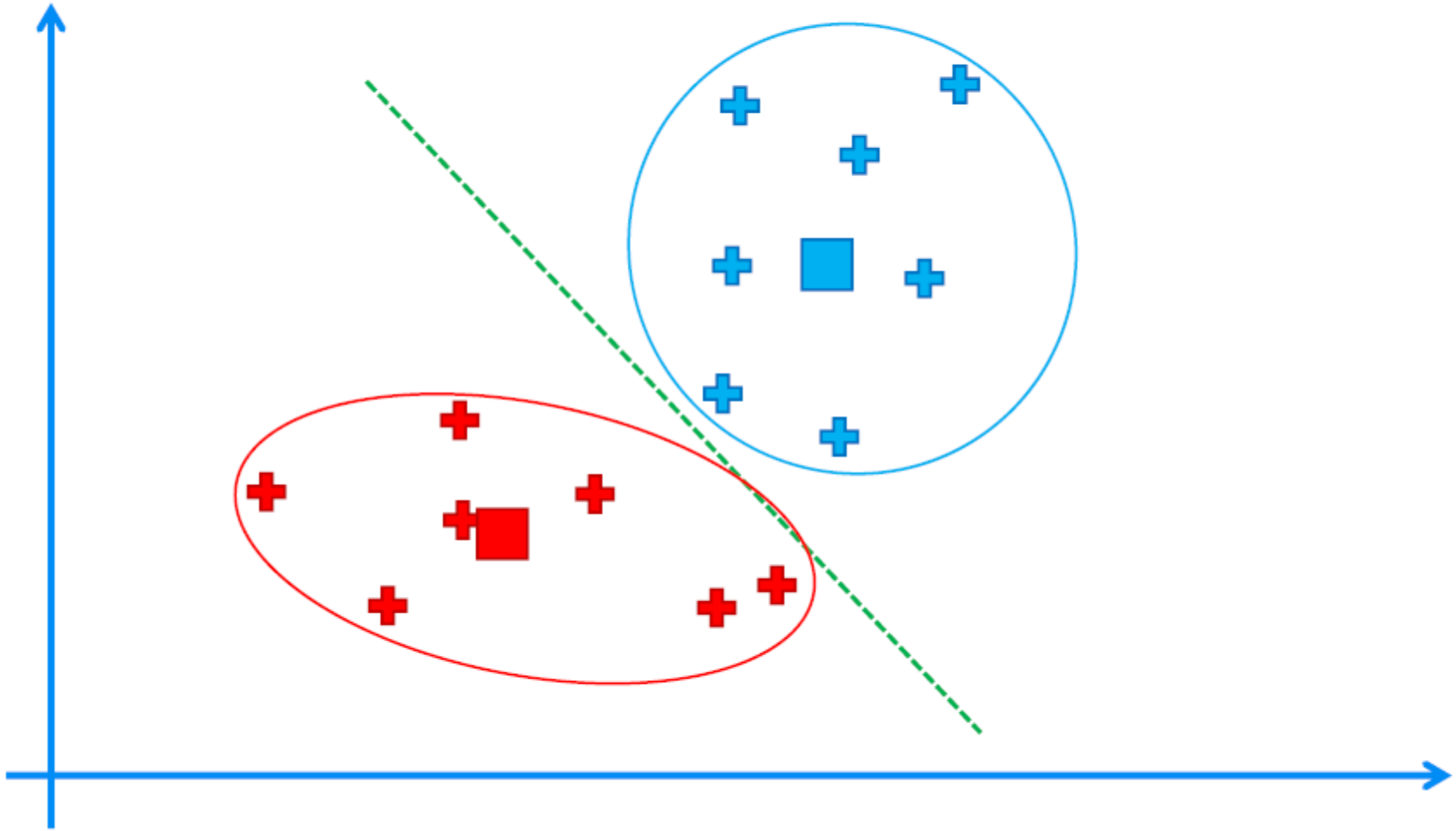
K means clustering algorithm



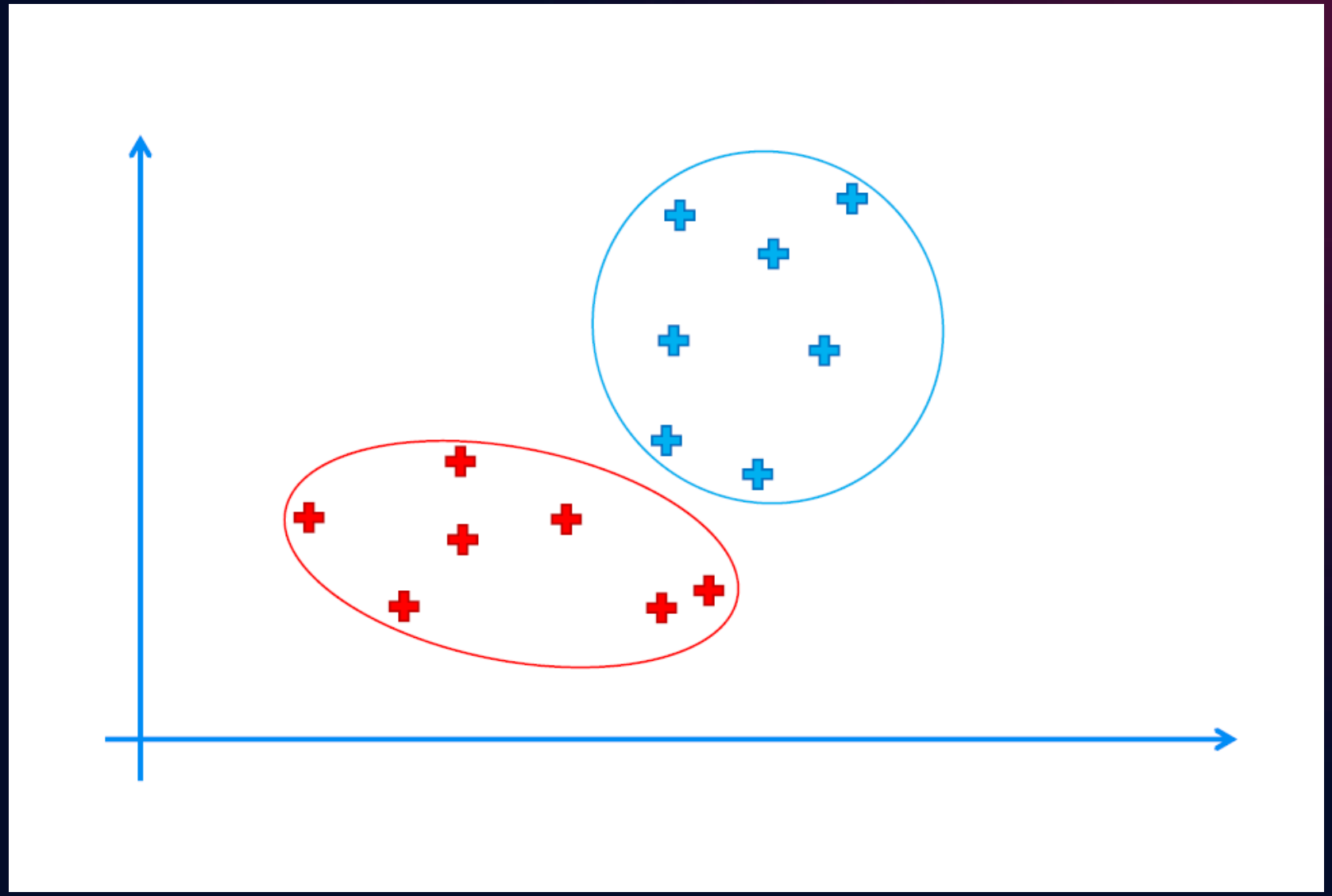
K means clustering algorithm



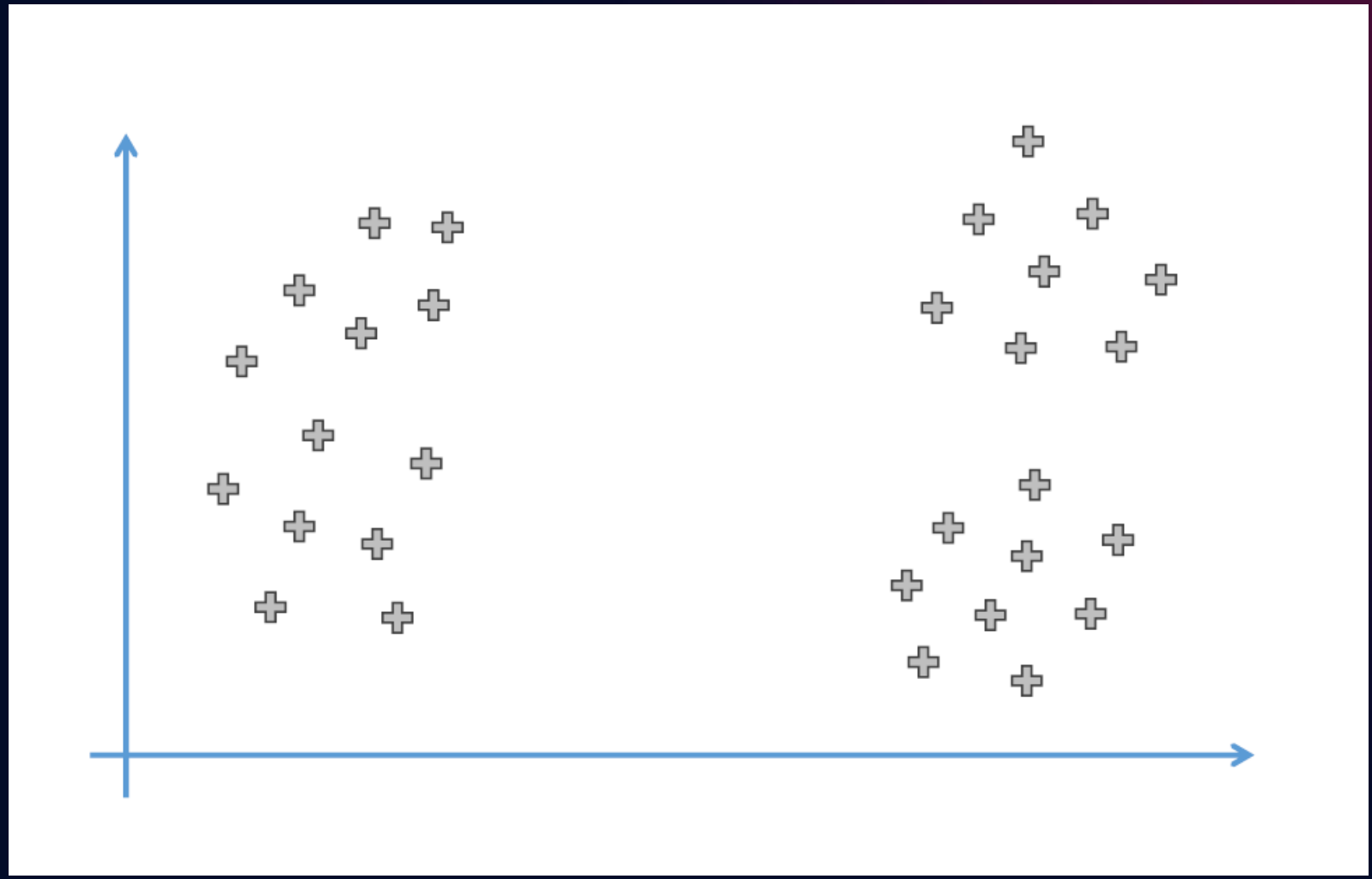
K means clustering algorithm



K means clustering algorithm



The Elbow Method

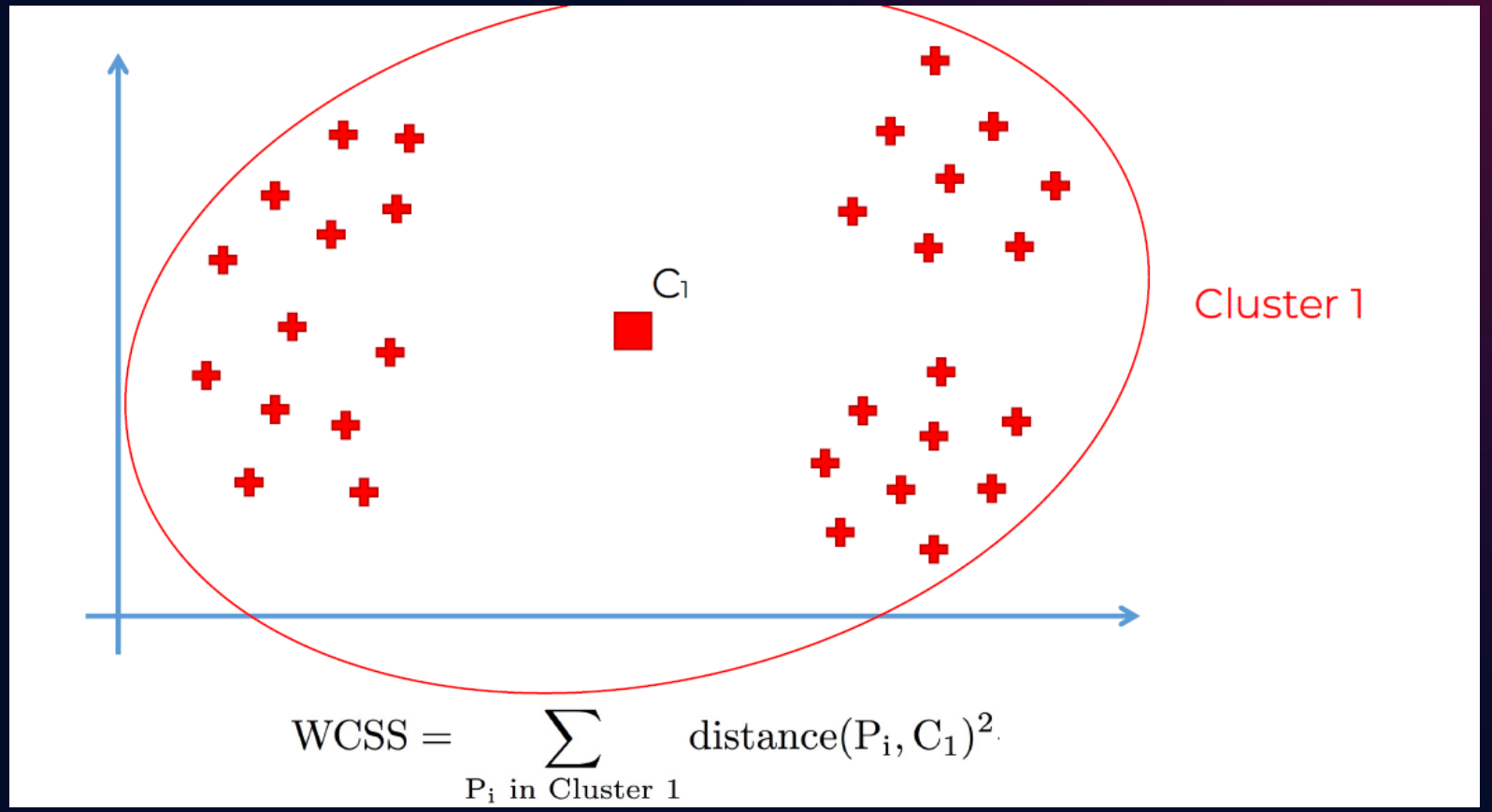


The Elbow Method

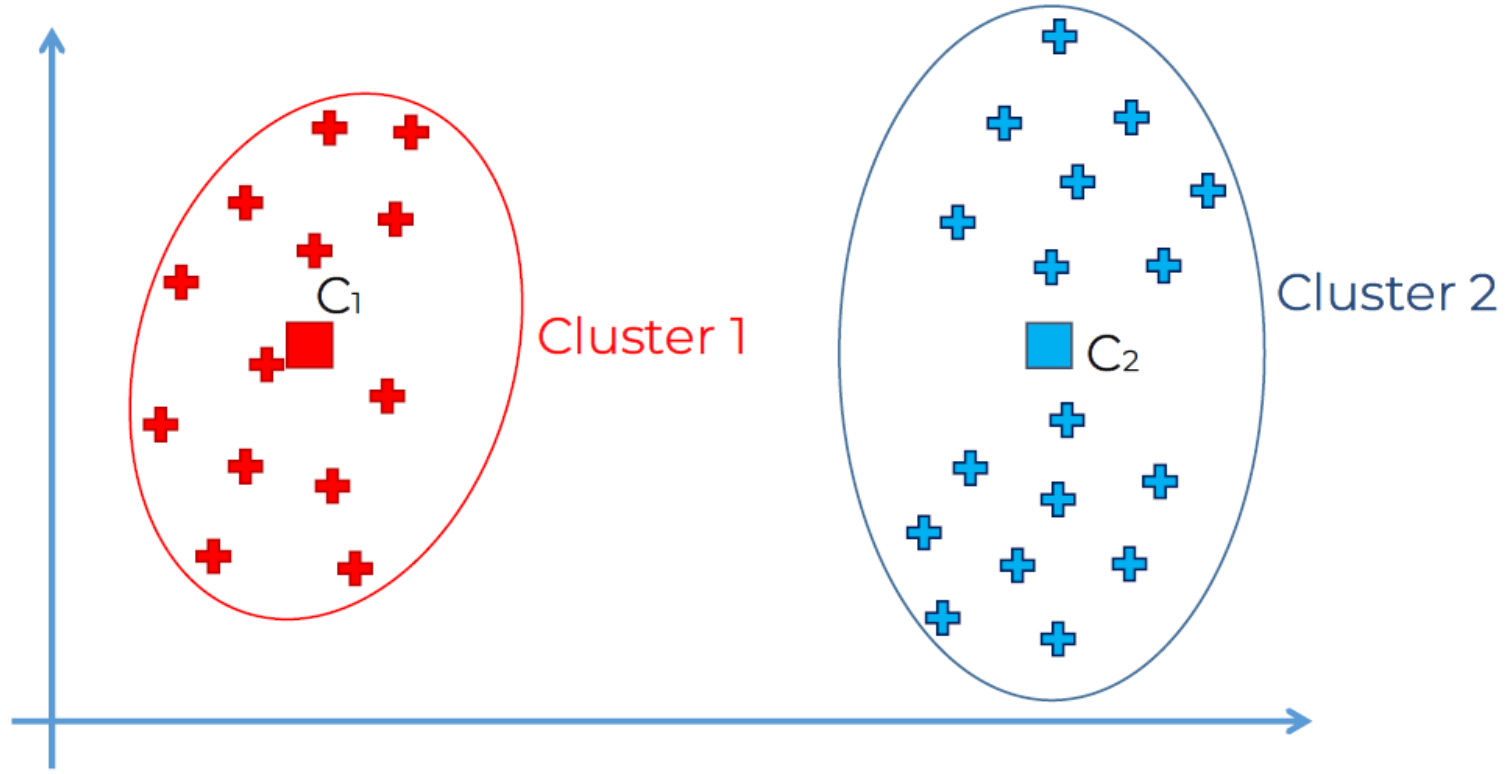
Within Cluster Sum of Squares:

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

The Elbow Method

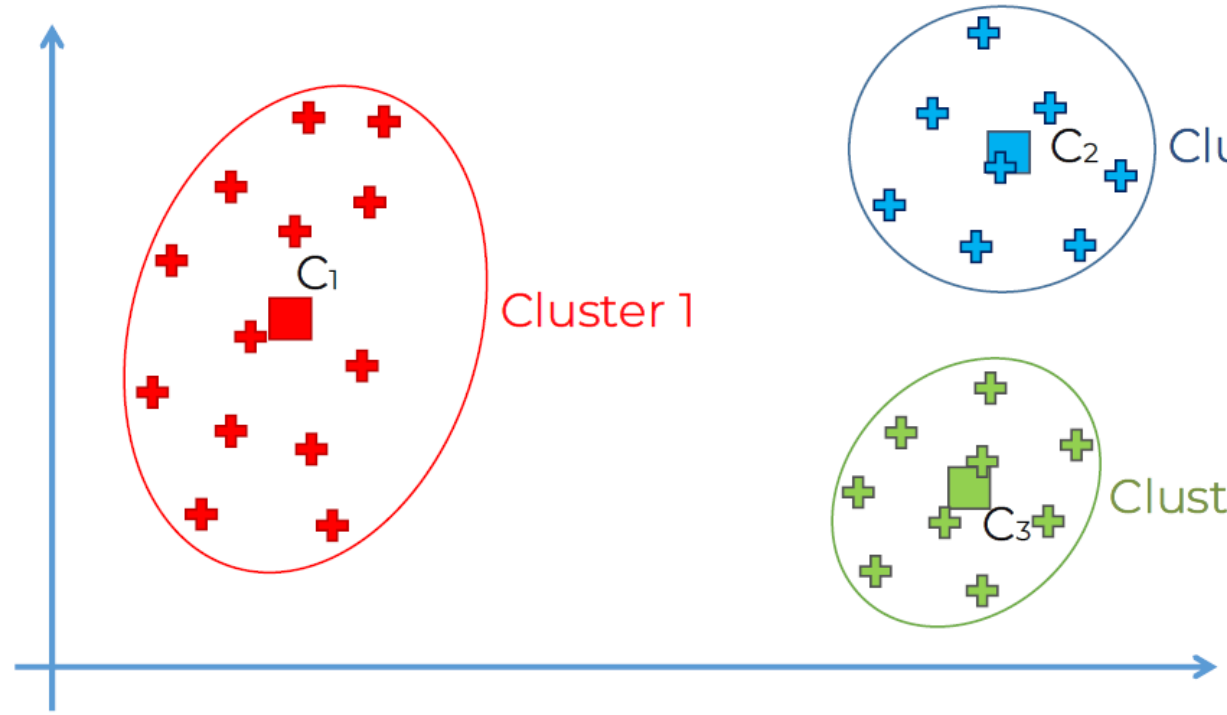


The Elbow Method



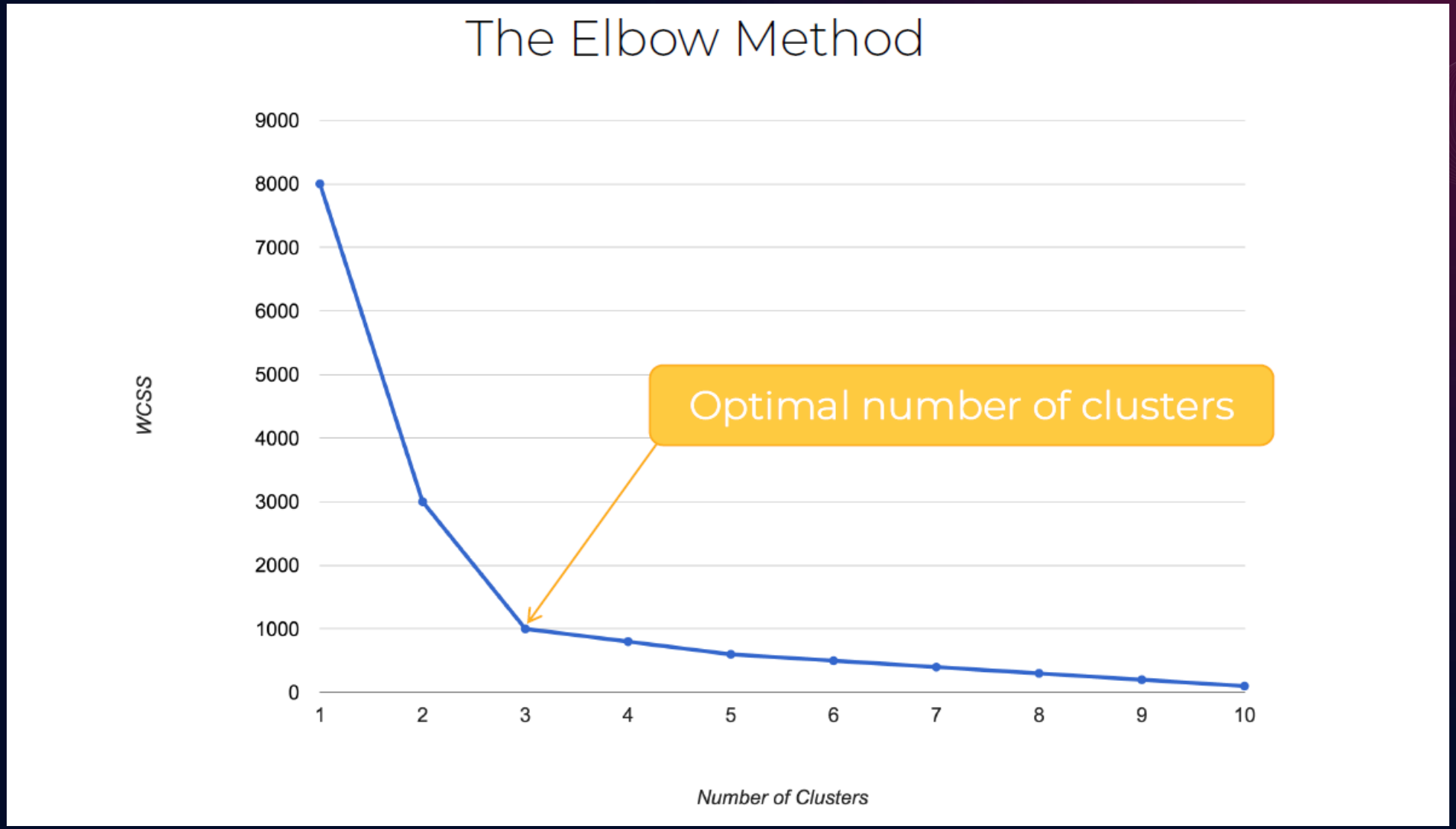
$$\text{WCSS} = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2$$

The Elbow Method

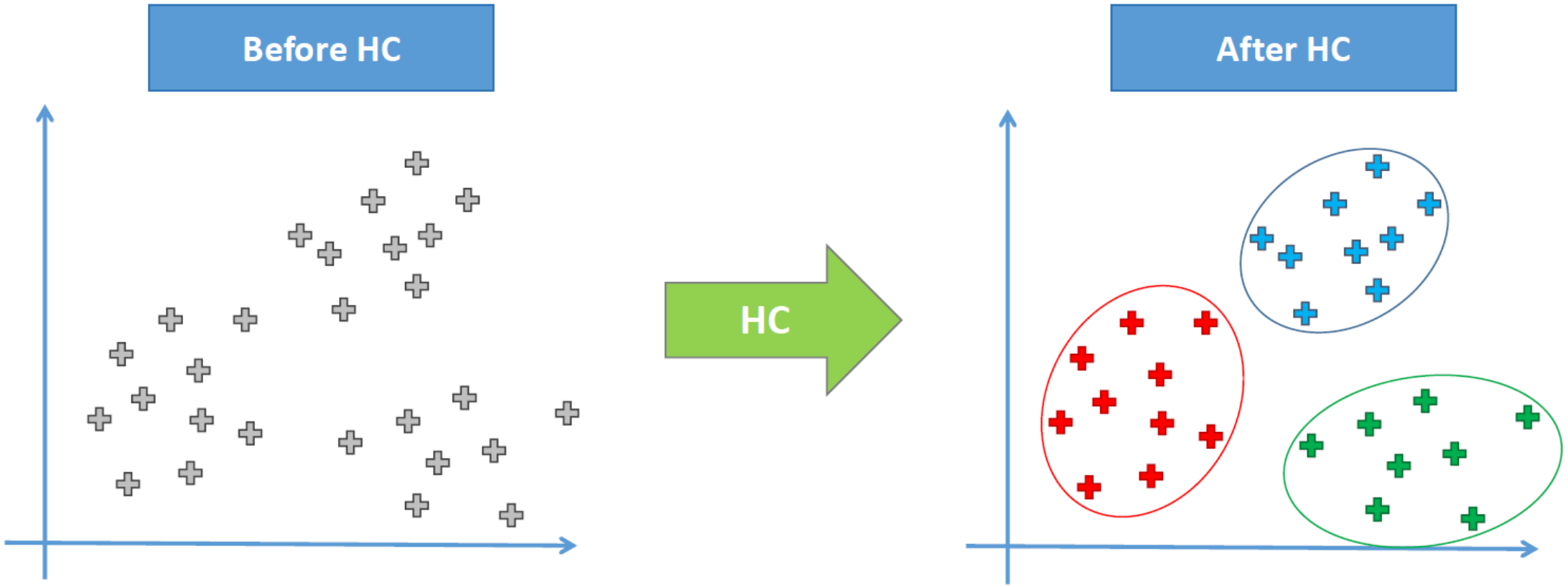


$$\text{WCSS} = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2 + \sum_{P_i \text{ in Cluster 3}} \text{distance}(P_i, C_3)^2$$

The Elbow Method



HC

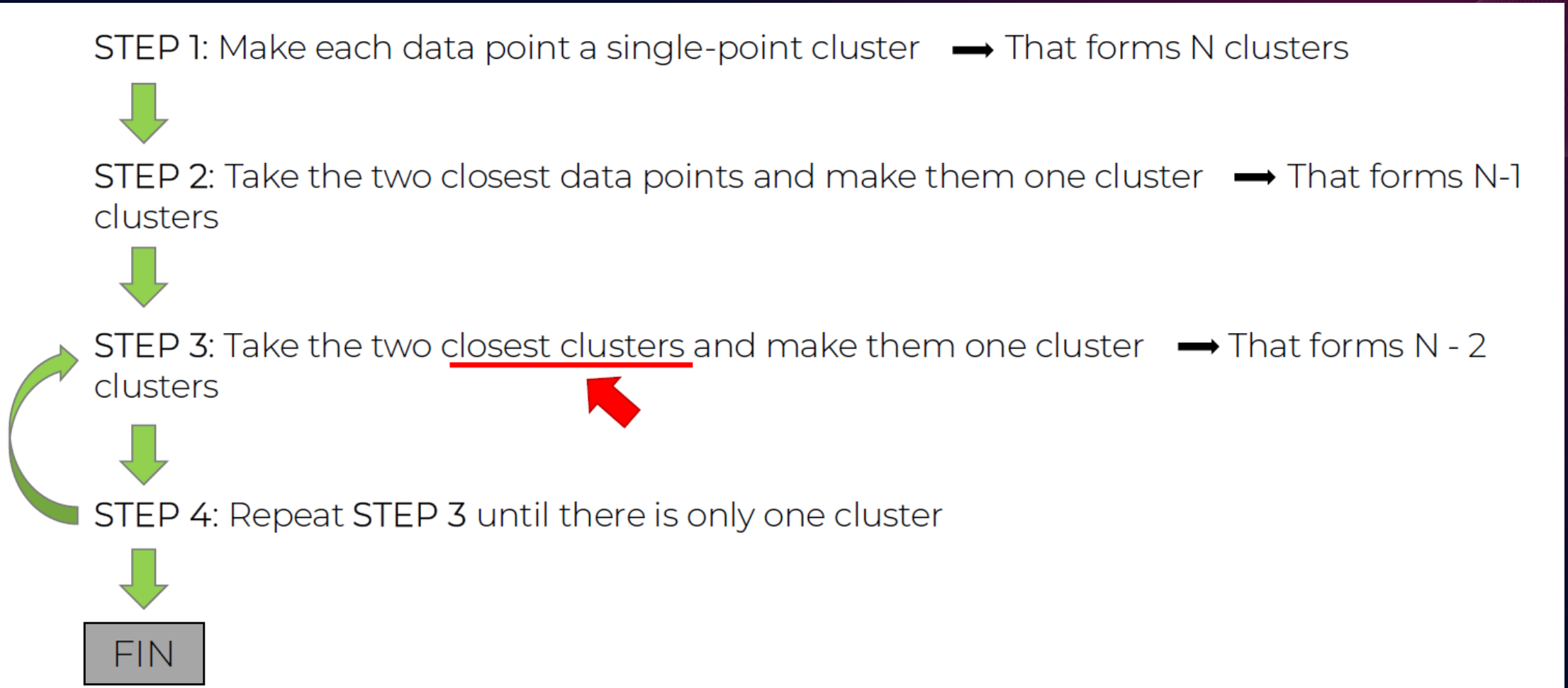


Same as K-Means but different process

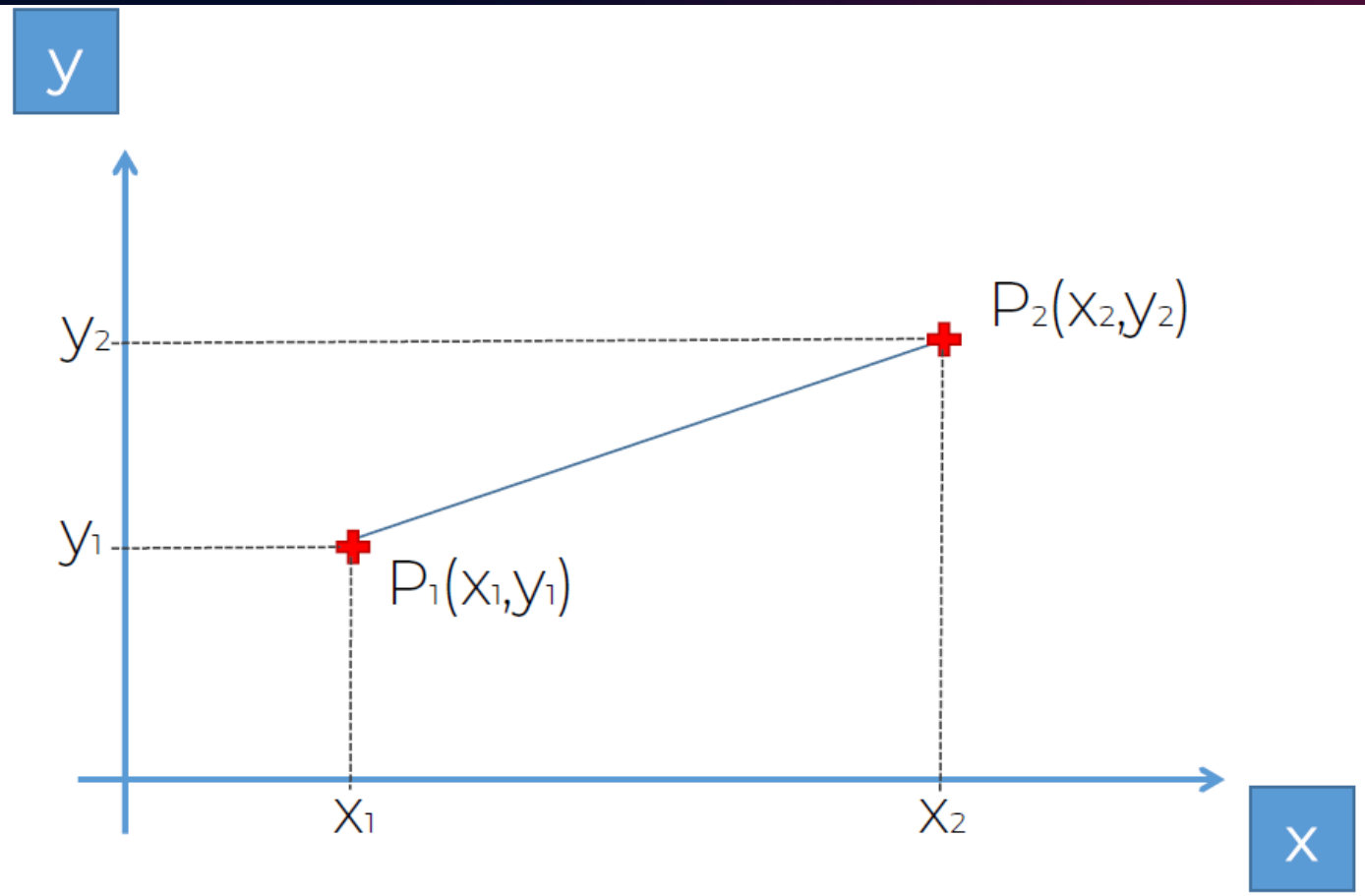
Hierarchical Clustering

NOTE:
Agglomerative
&
Divisive

Agglomerative HC



Agglomerative HC

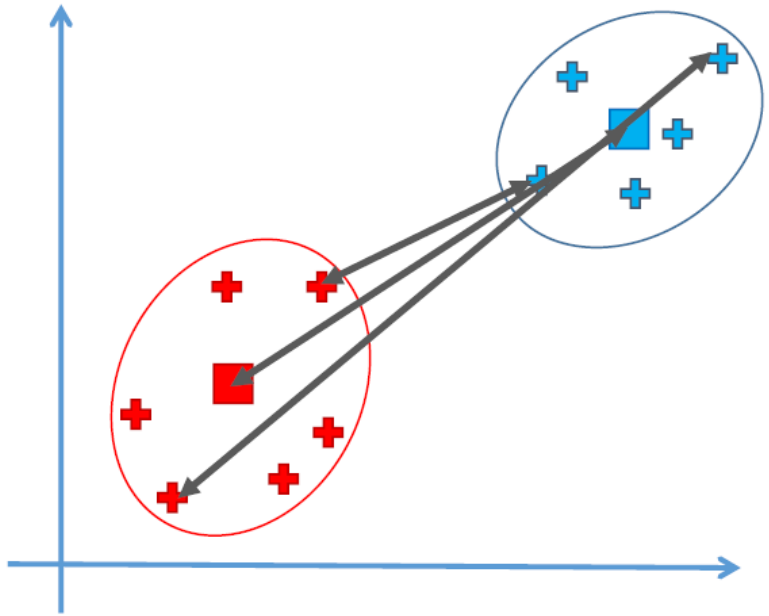


$$\text{Euclidean Distance between } P_1 \text{ and } P_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Closest pair of clusters

- Many variants to defining closest pair of clusters
- **Single-link**
  - Similarity of the *most* cosine-similar (single-link)
- **Complete-link**
  - Similarity of the "furthest" points, the *least* cosine-similar
- **Centroid**
  - Clusters whose centroids (centers of gravity) are the most cosine-similar
- **Average-link**
  - Average cosine between pairs of elements

Agglomerative HC



Distance Between Two Clusters:

- Option 1: Closest Points
- Option 2: Furthest Points
- Option 3: Average Distance
- Option 4: Distance Between Centroids

Agglomerative HC

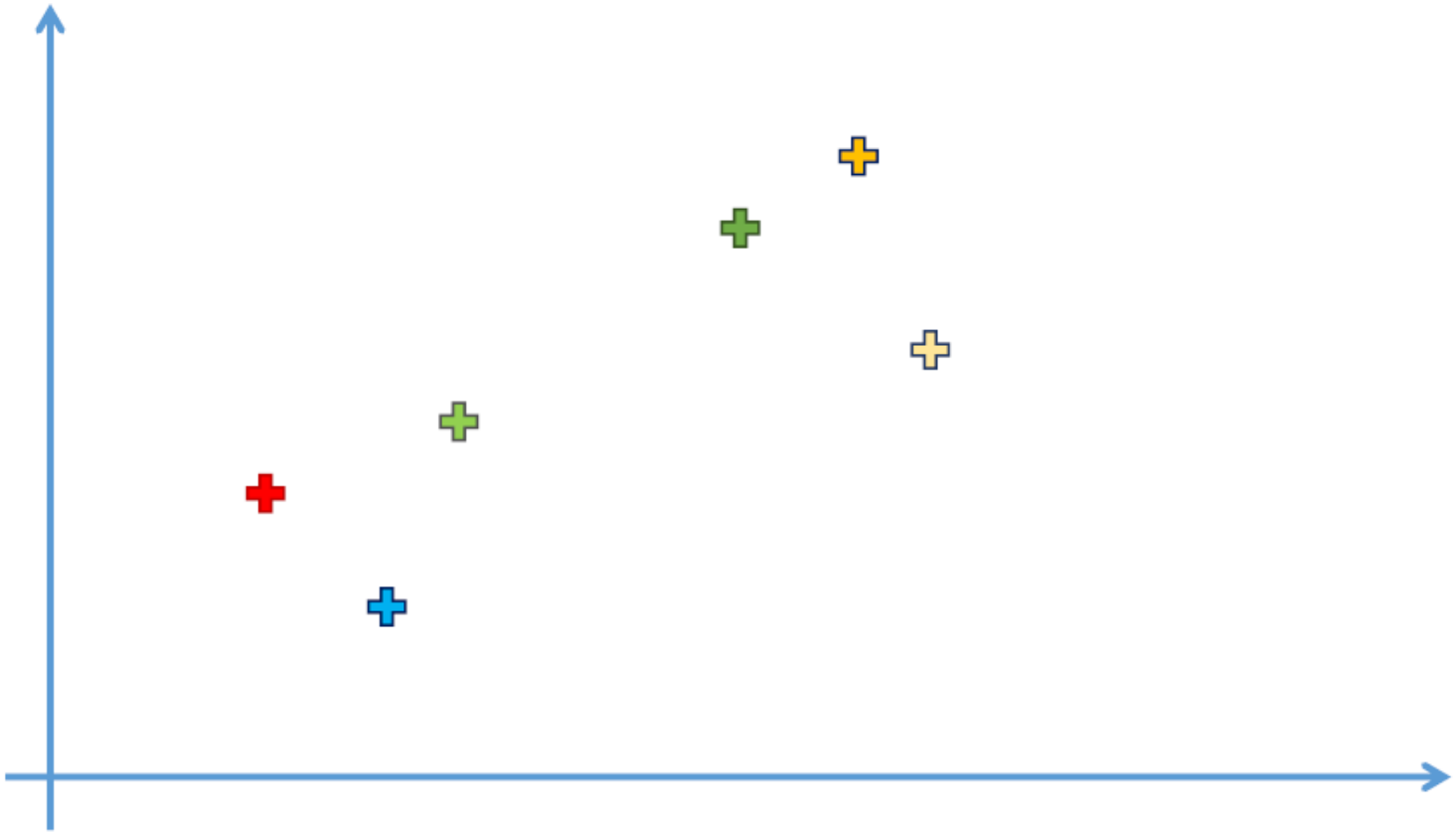
Consider the following dataset of N = 6 data points

Agglomerative HC

STEP 1: Make each data point a single-point cluster ➞ That forms 6 clusters

Agglomerative HC

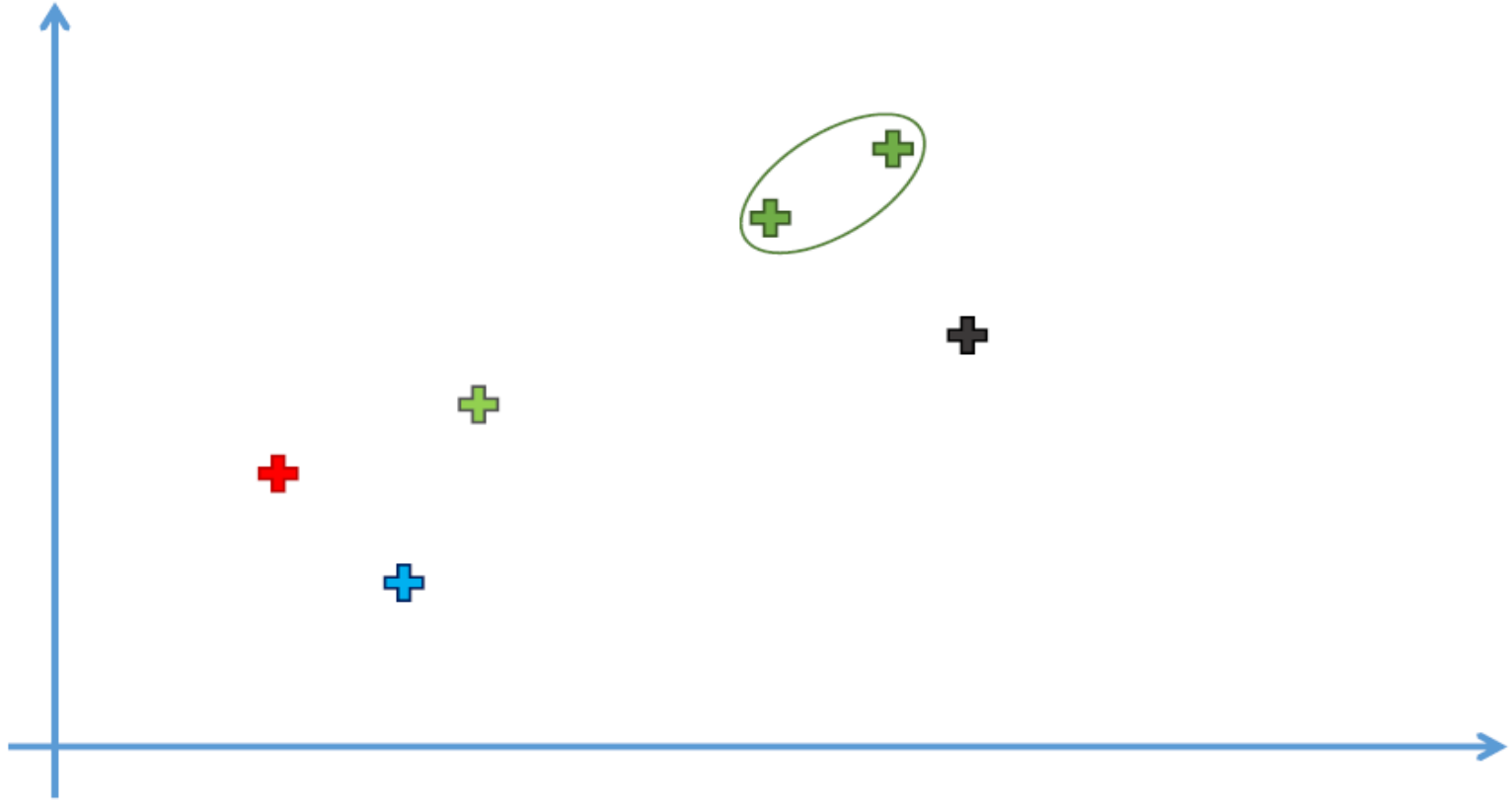
STEP 1: Make each data point a single-point cluster → That forms 6 clusters



Agglomerative HC

STEP 2: Take the two closest data points and make them one cluster

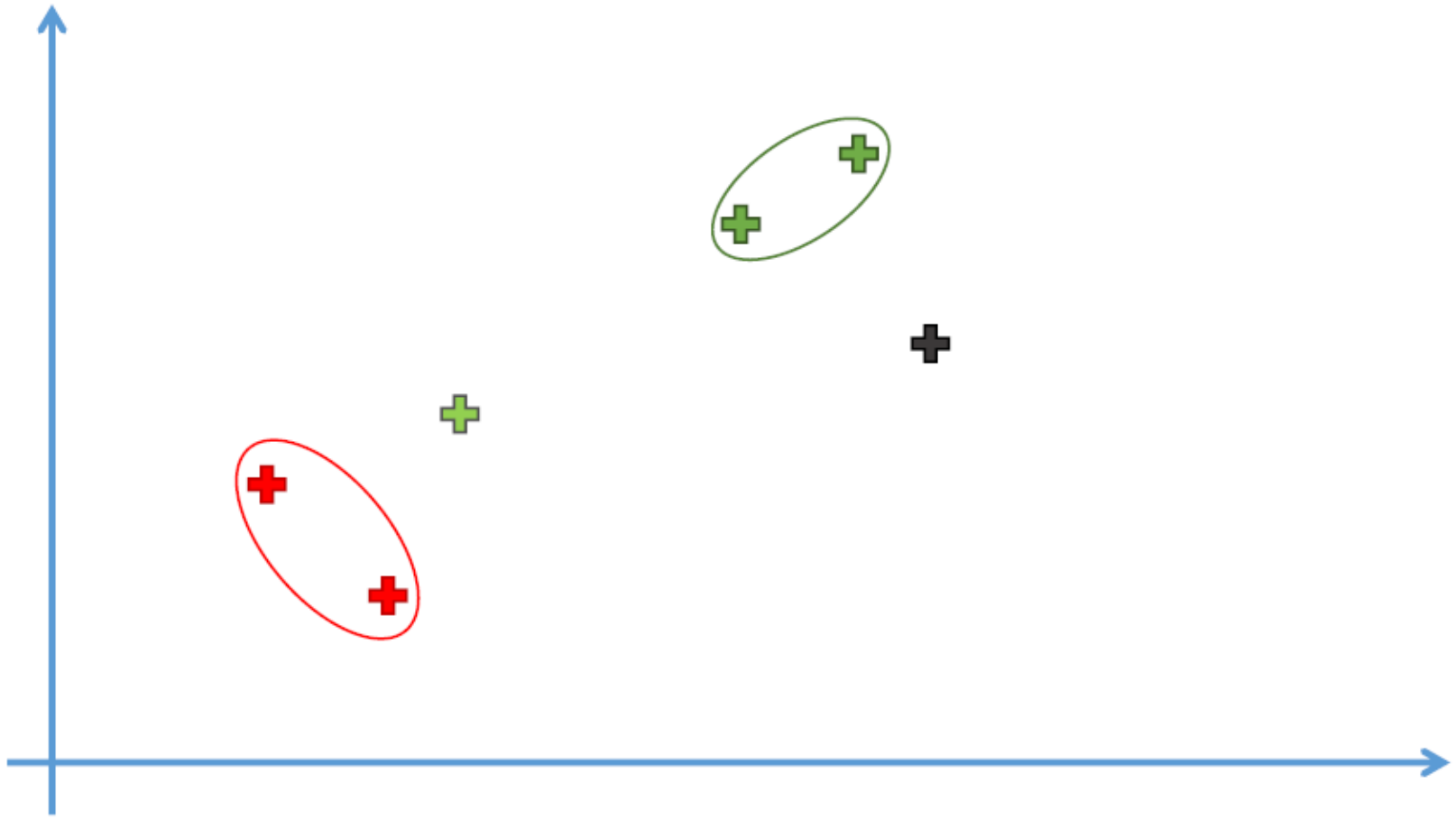
→ That forms 5 clusters



Agglomerative HC

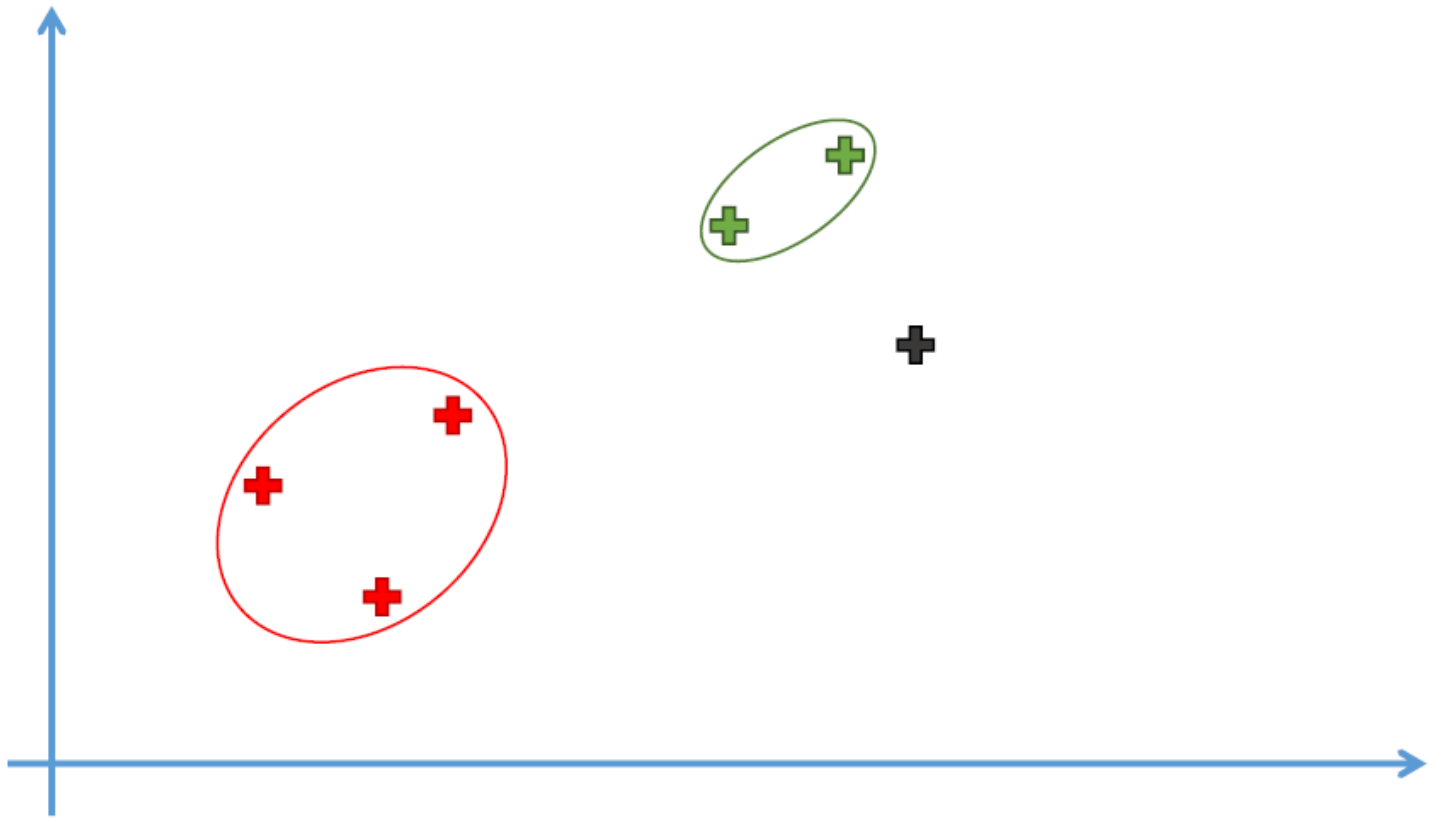
STEP 3: Take the two closest clusters and make them one cluster

→That forms 4 clusters



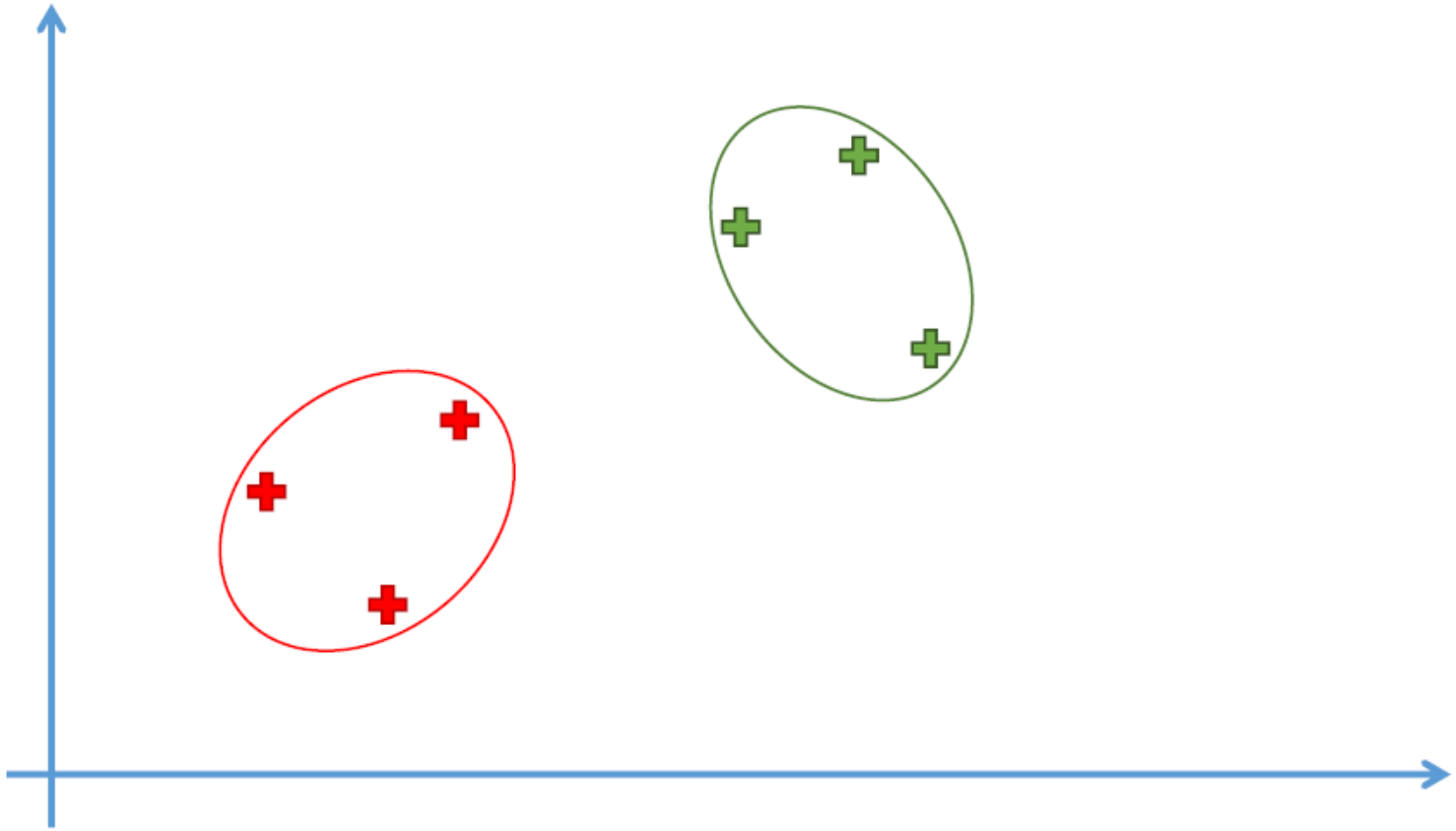
Agglomerative HC

STEP 4: Repeat STEP 3 until there is only one cluster



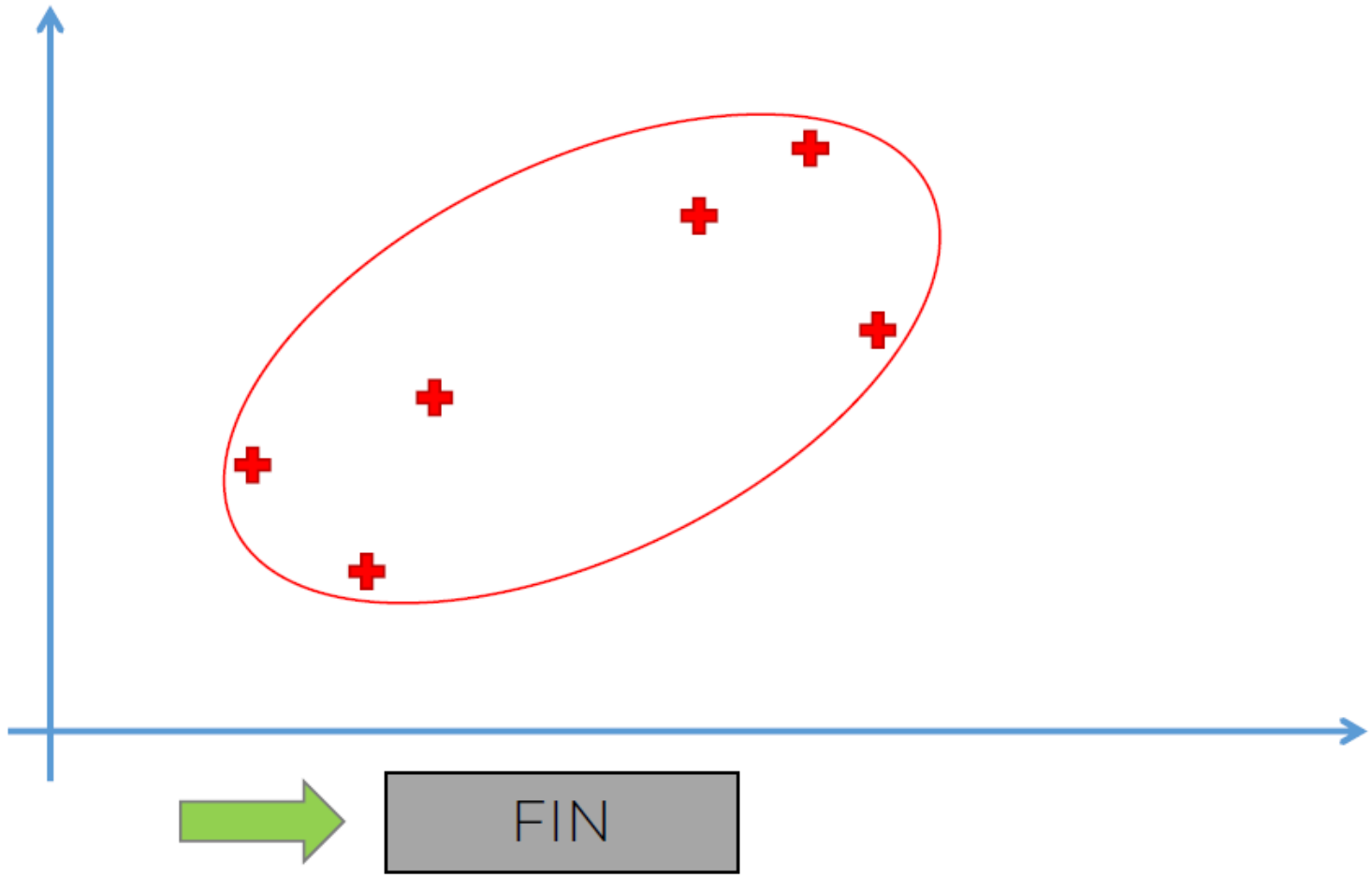
Agglomerative HC

STEP 4: Repeat STEP 3 until there is only one cluster

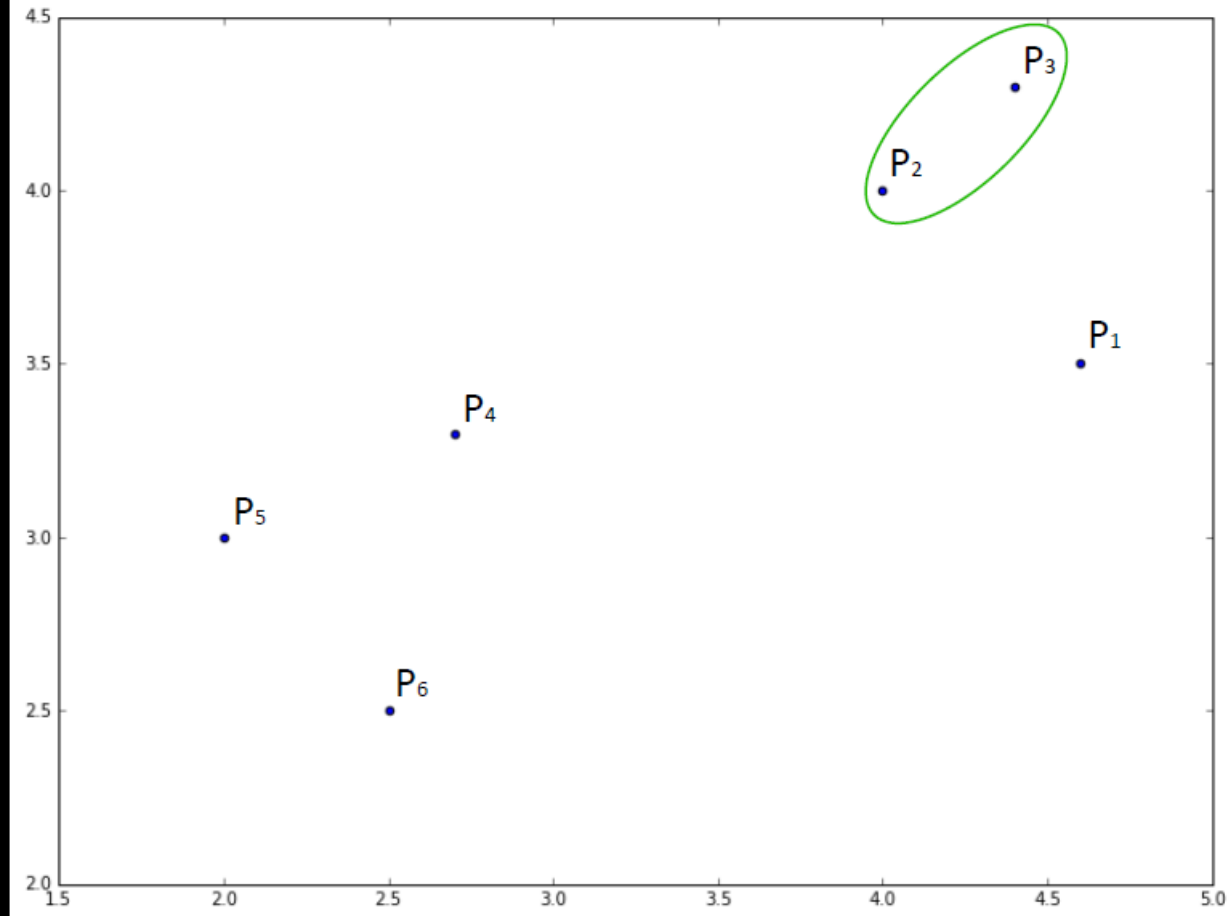


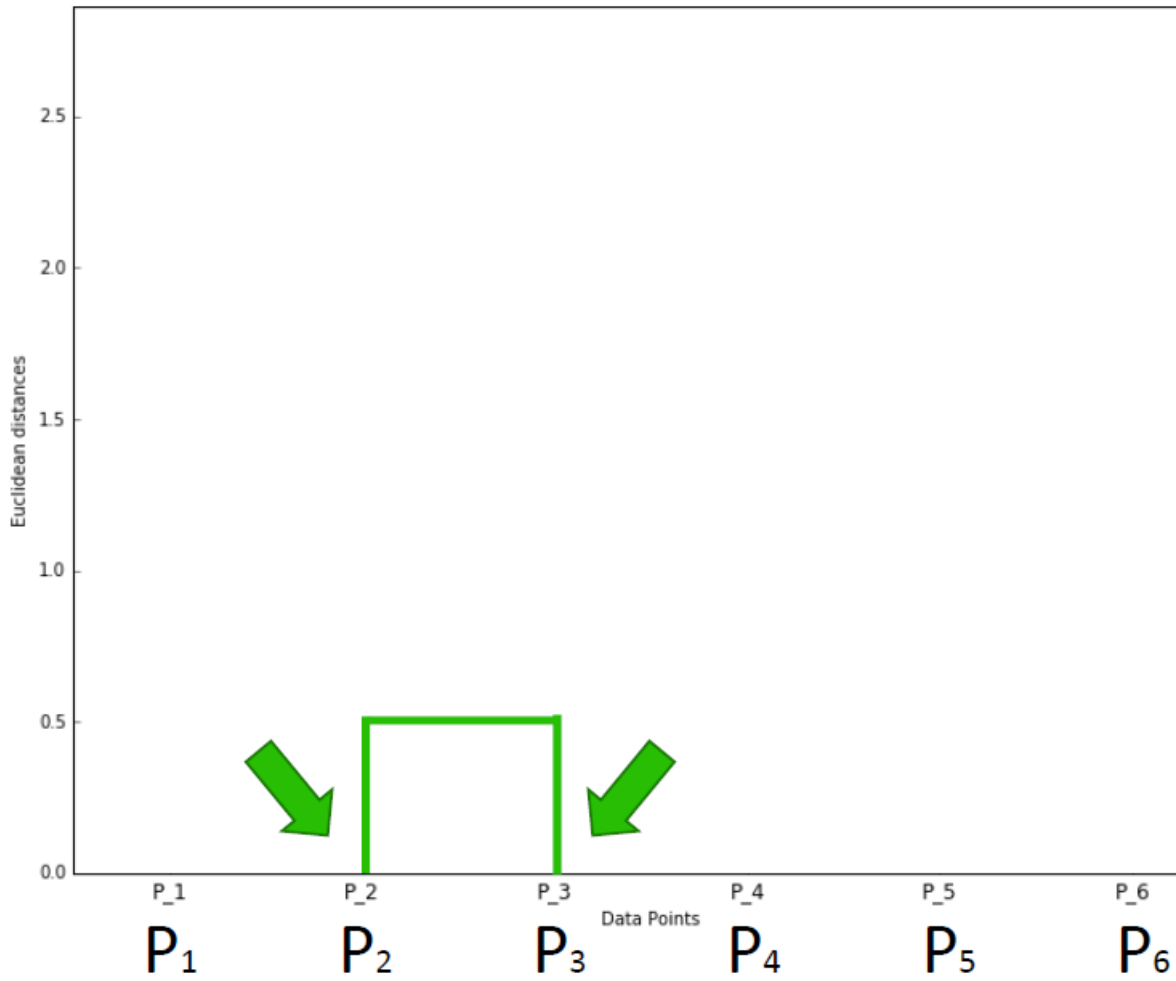
Agglomerative HC

STEP 4: Repeat STEP 3 until there is only one cluster

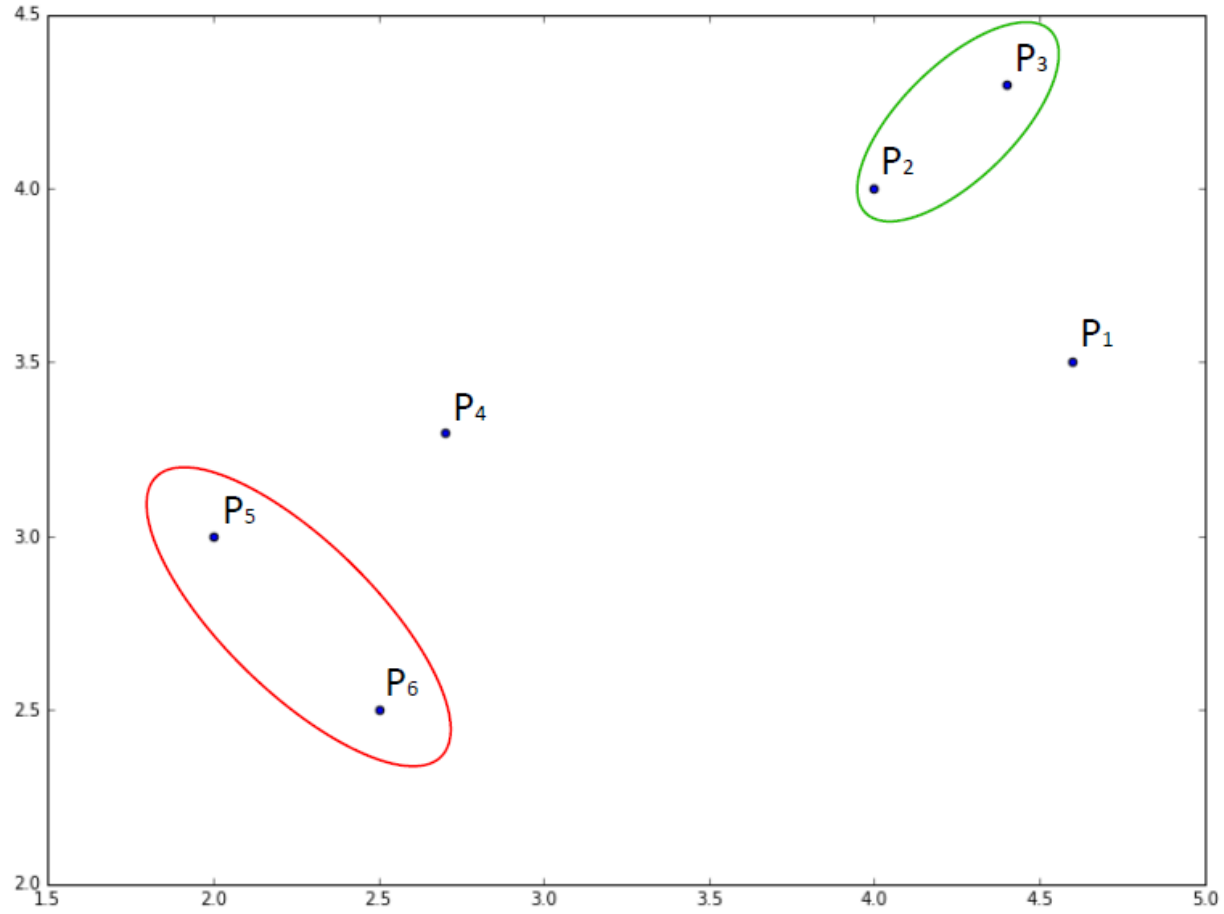


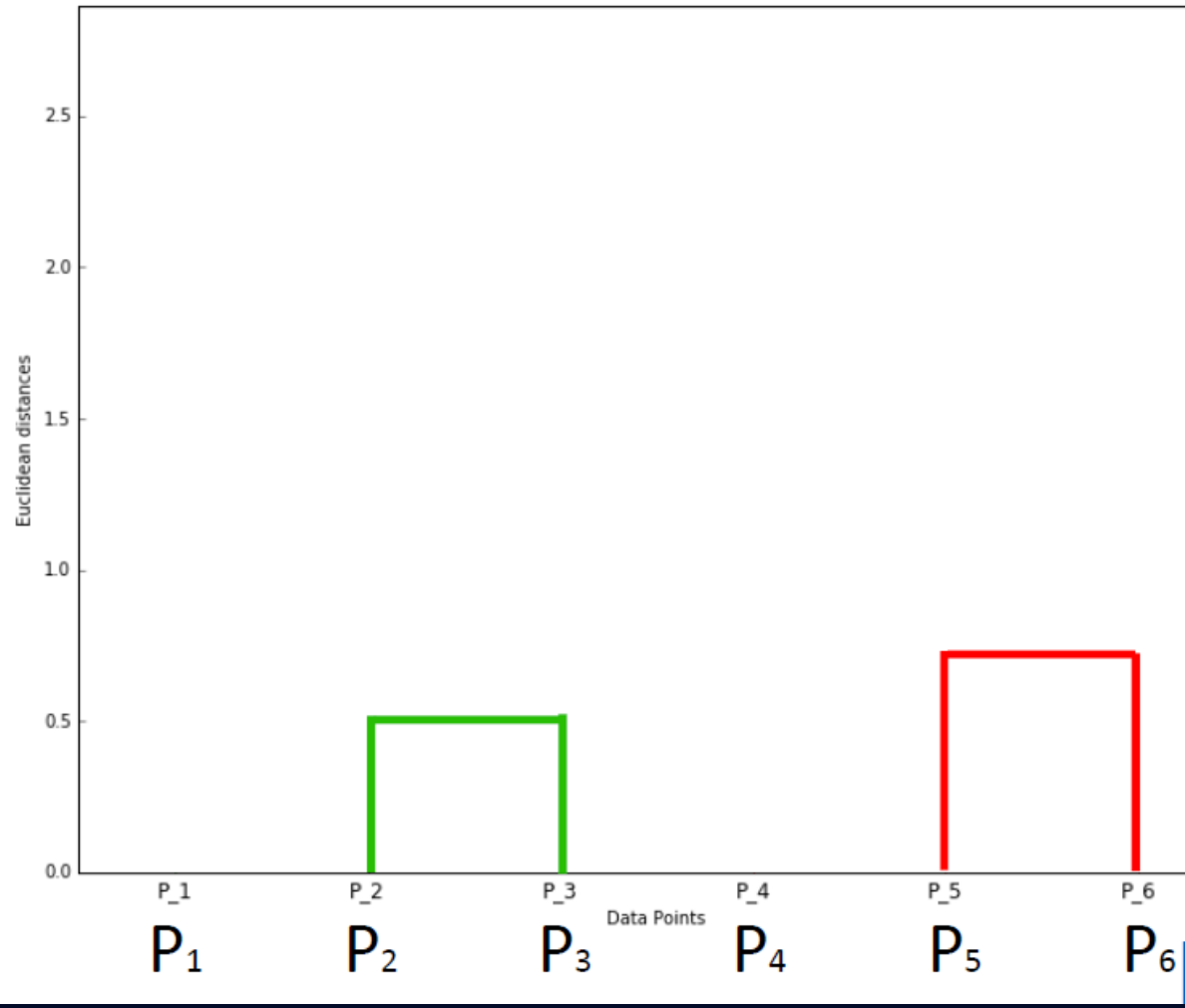
How Do Dendrograms works?



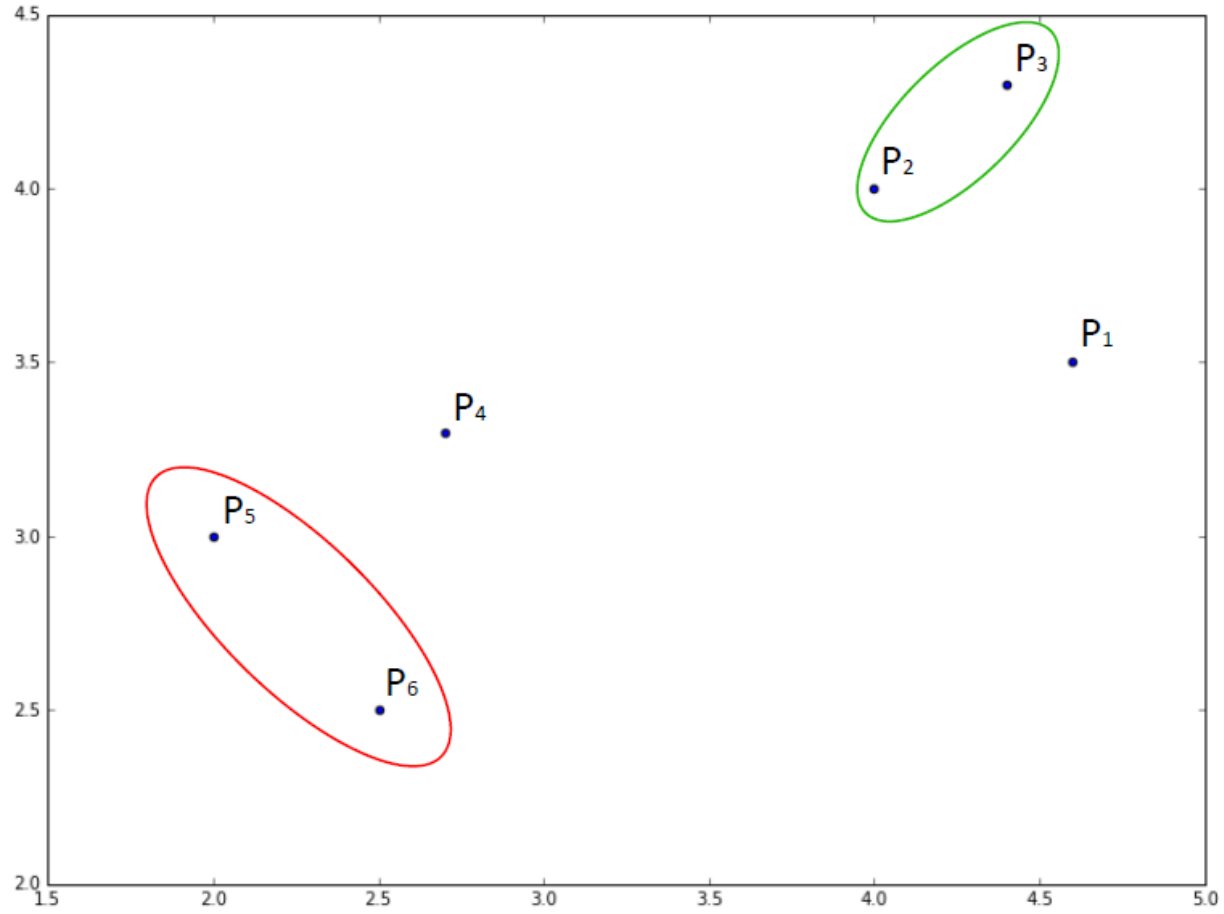


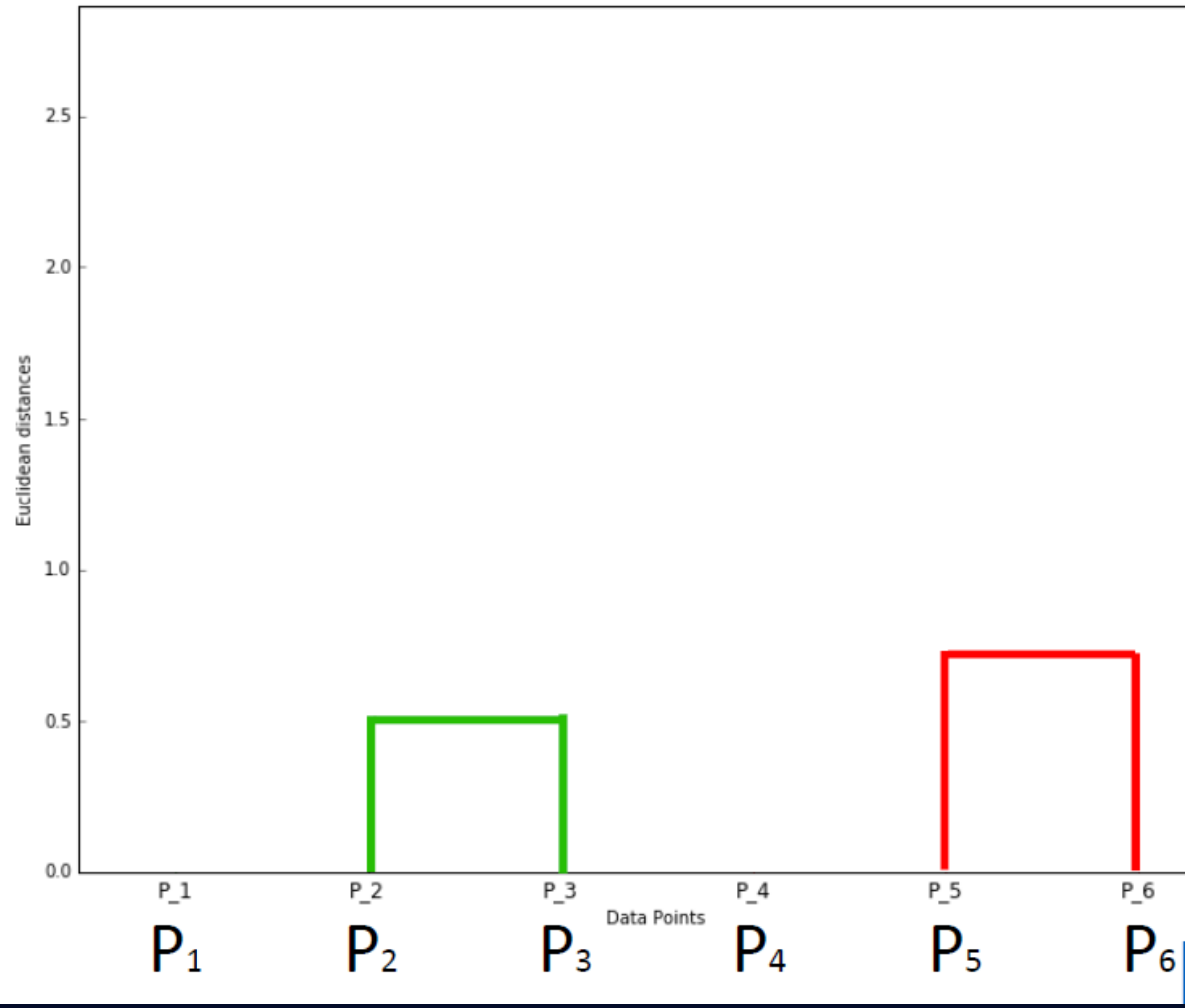
How Do Dendrograms works?



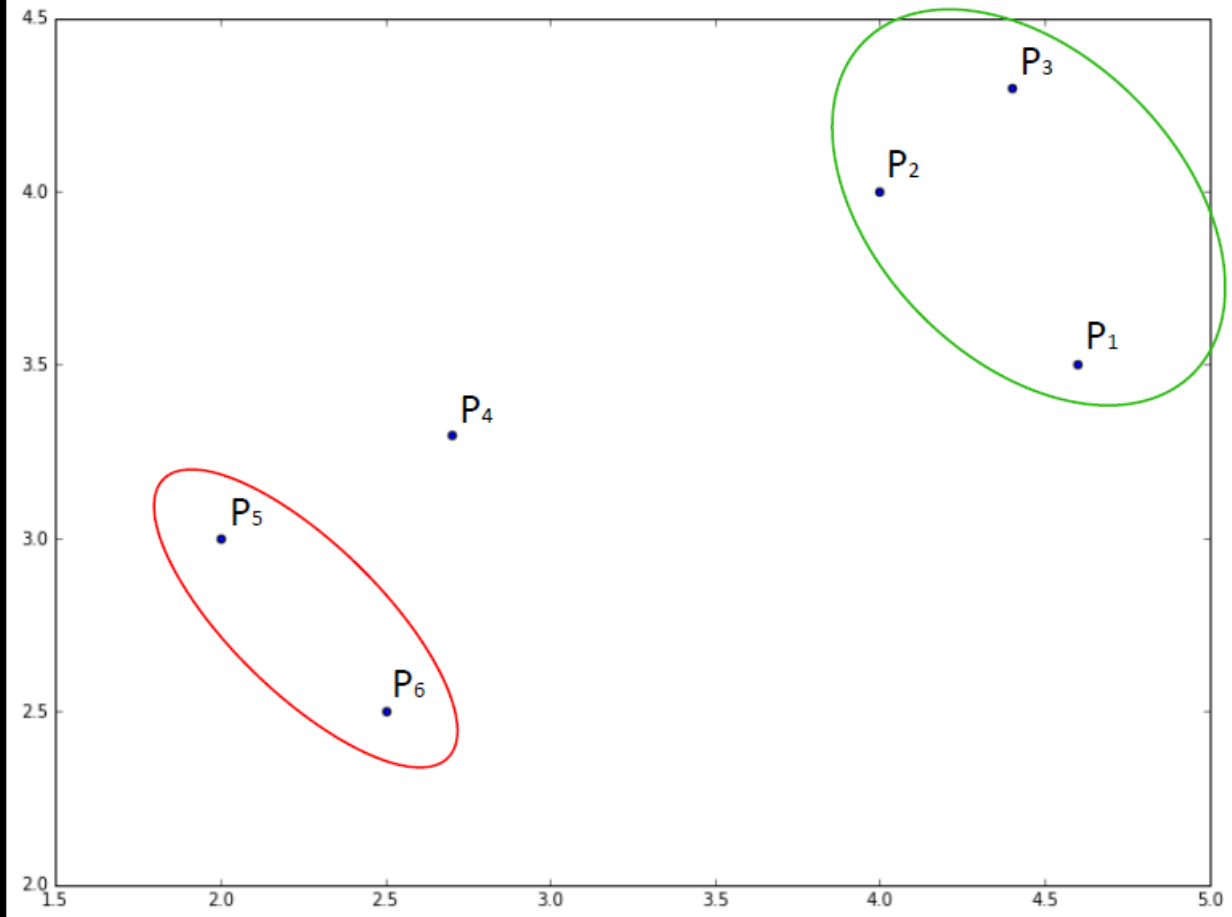


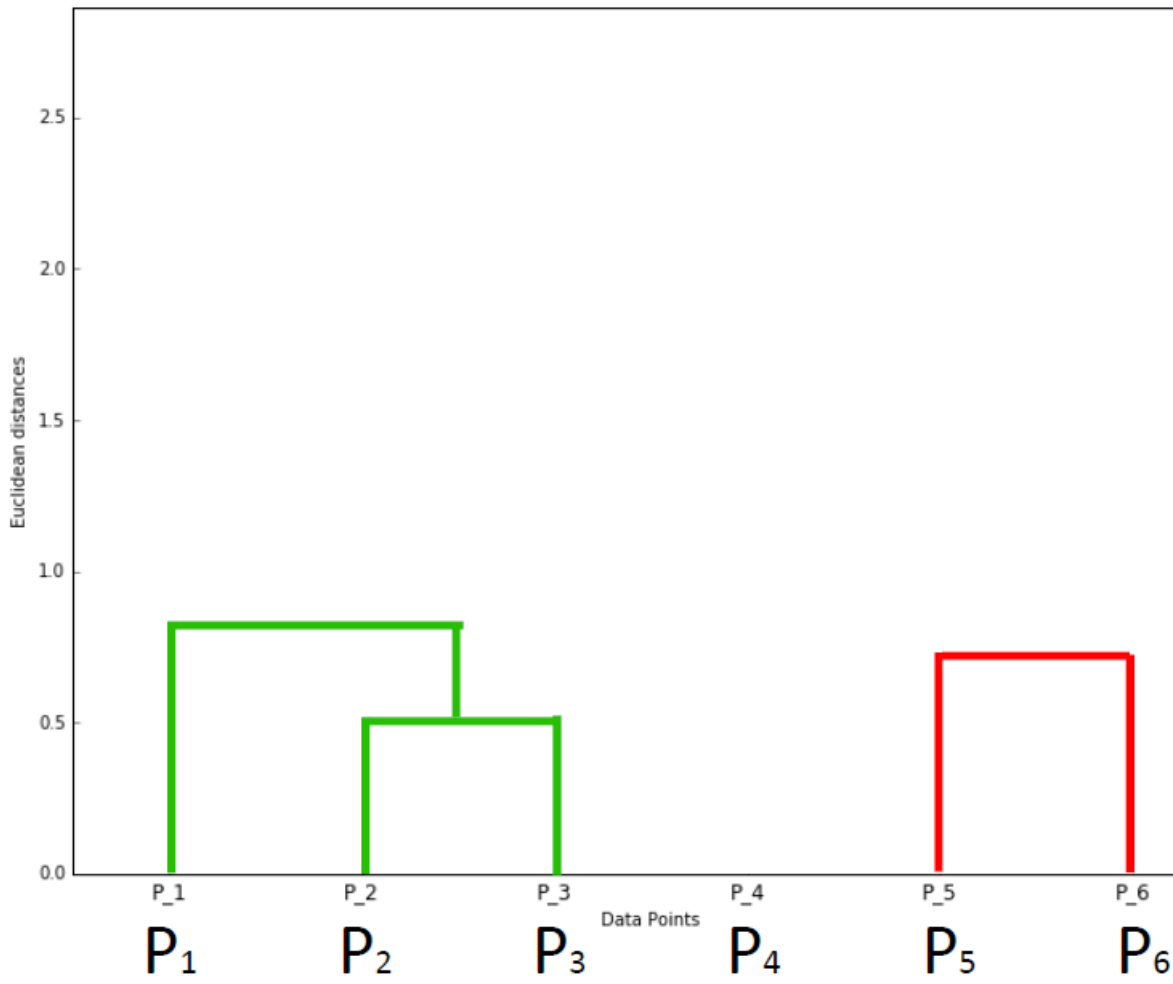
How Do Dendrograms works?



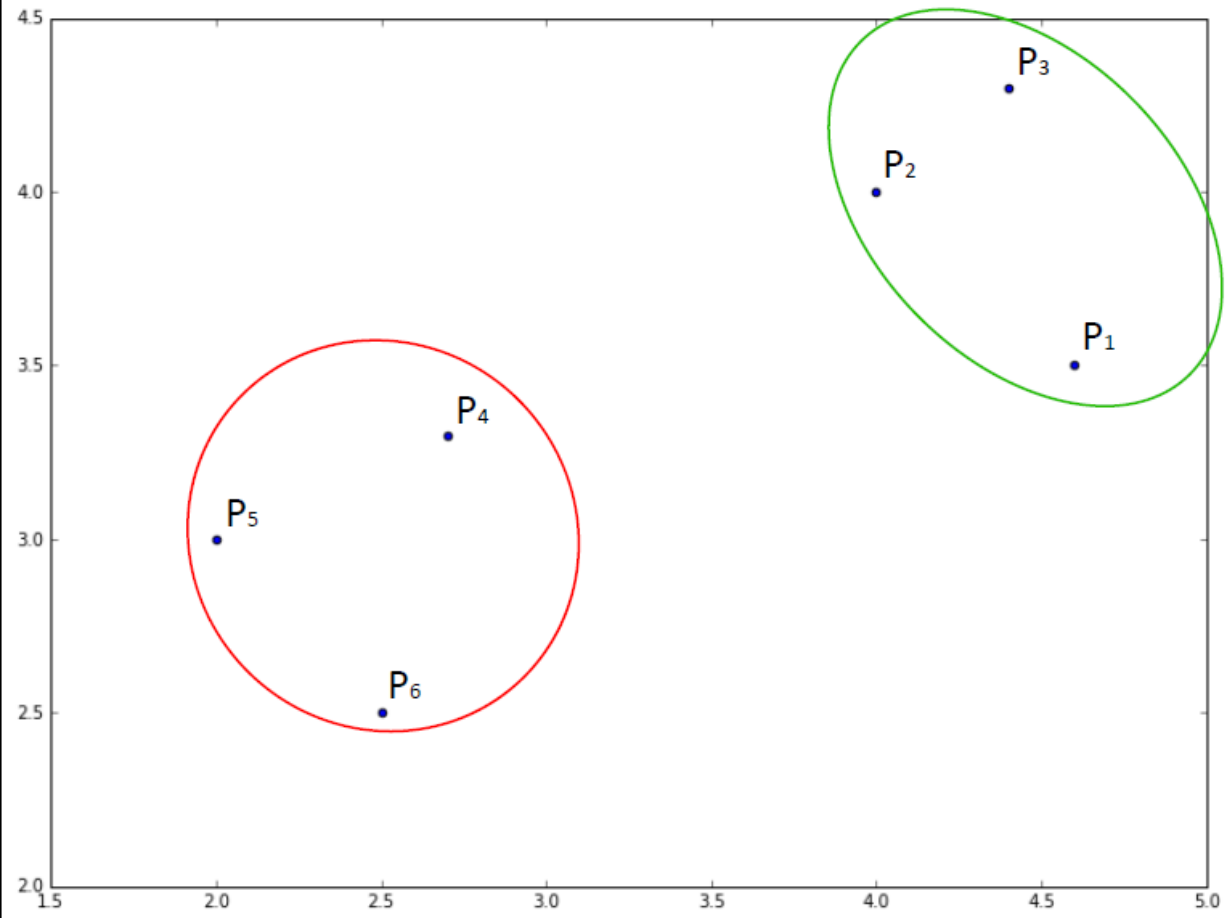


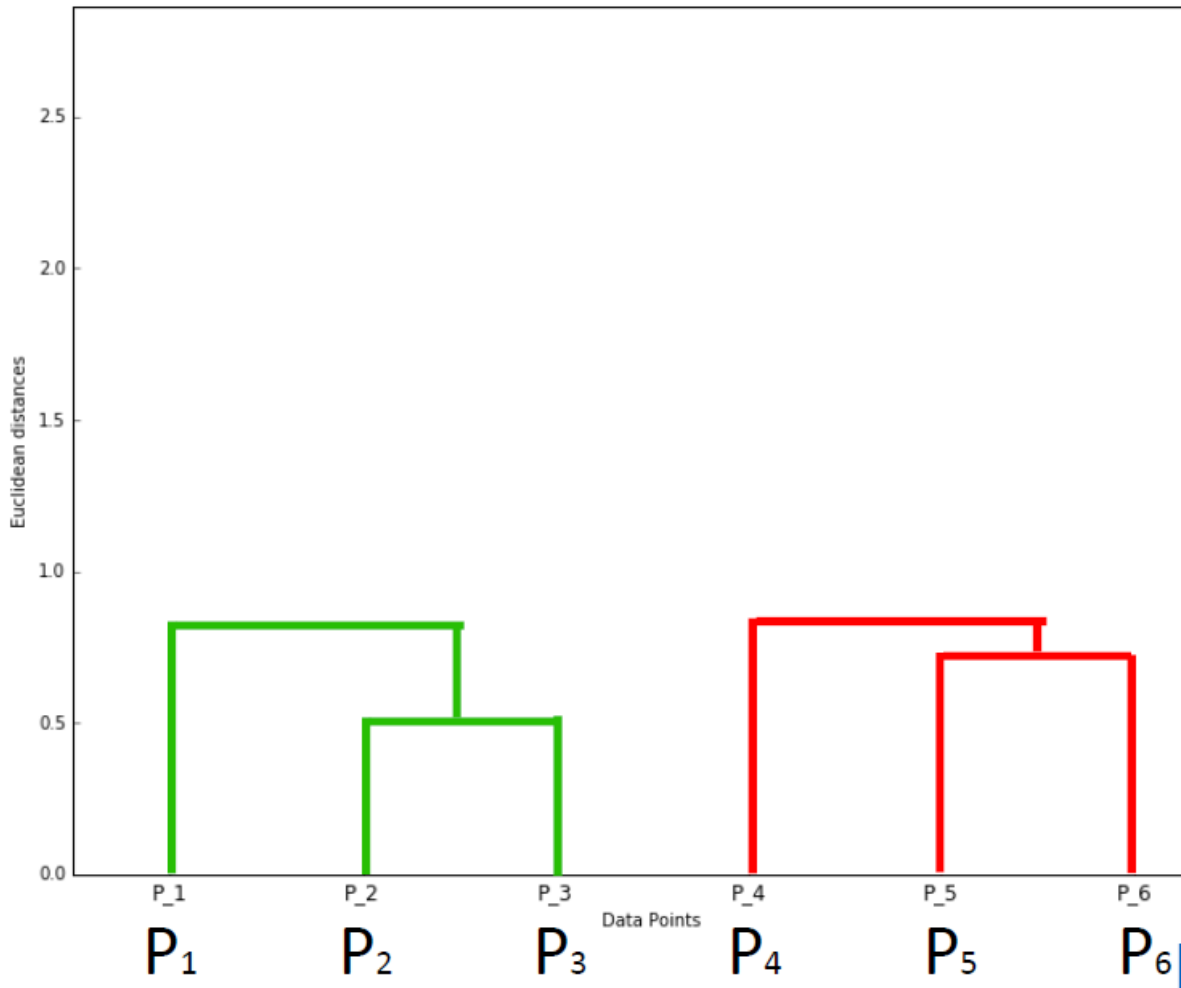
How Do Dendrograms works?



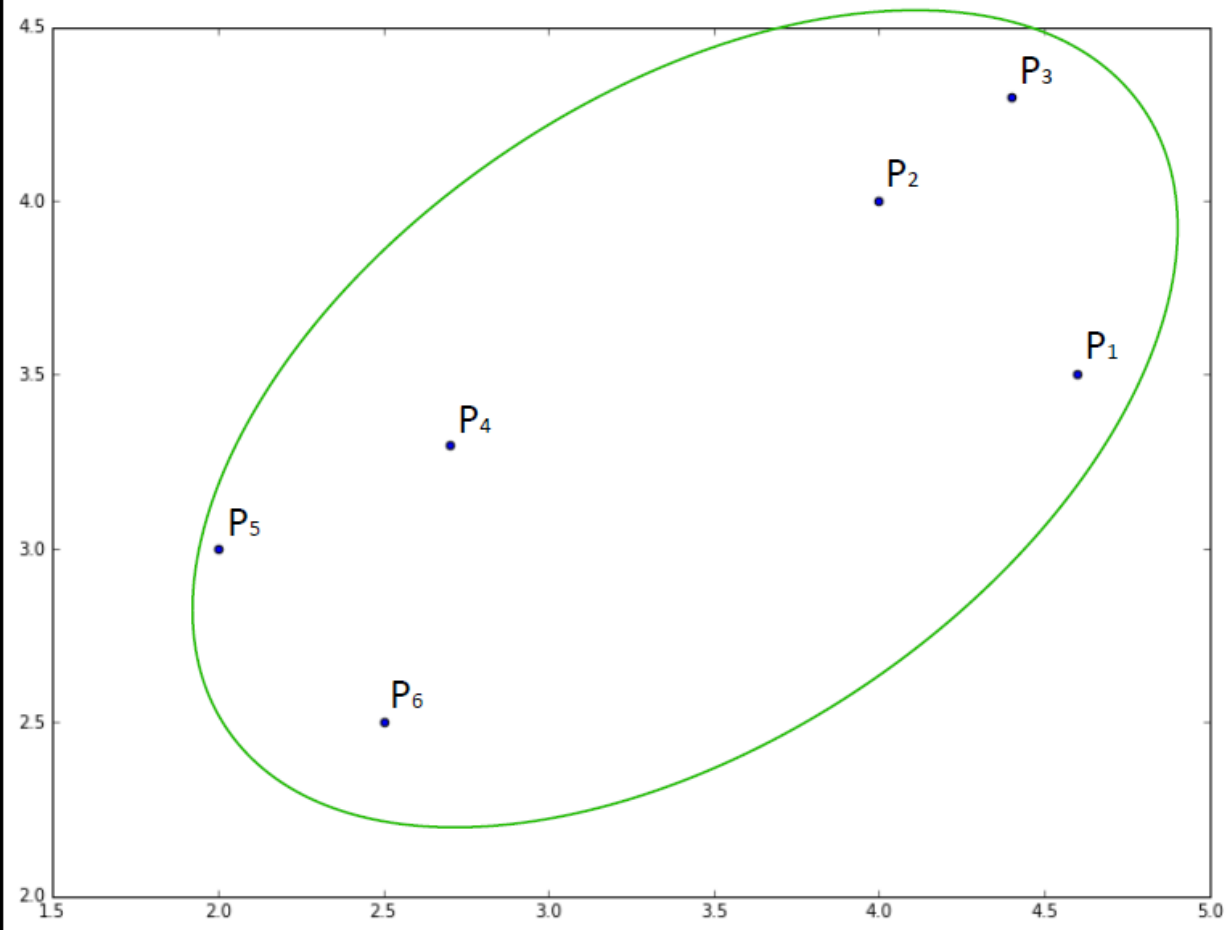


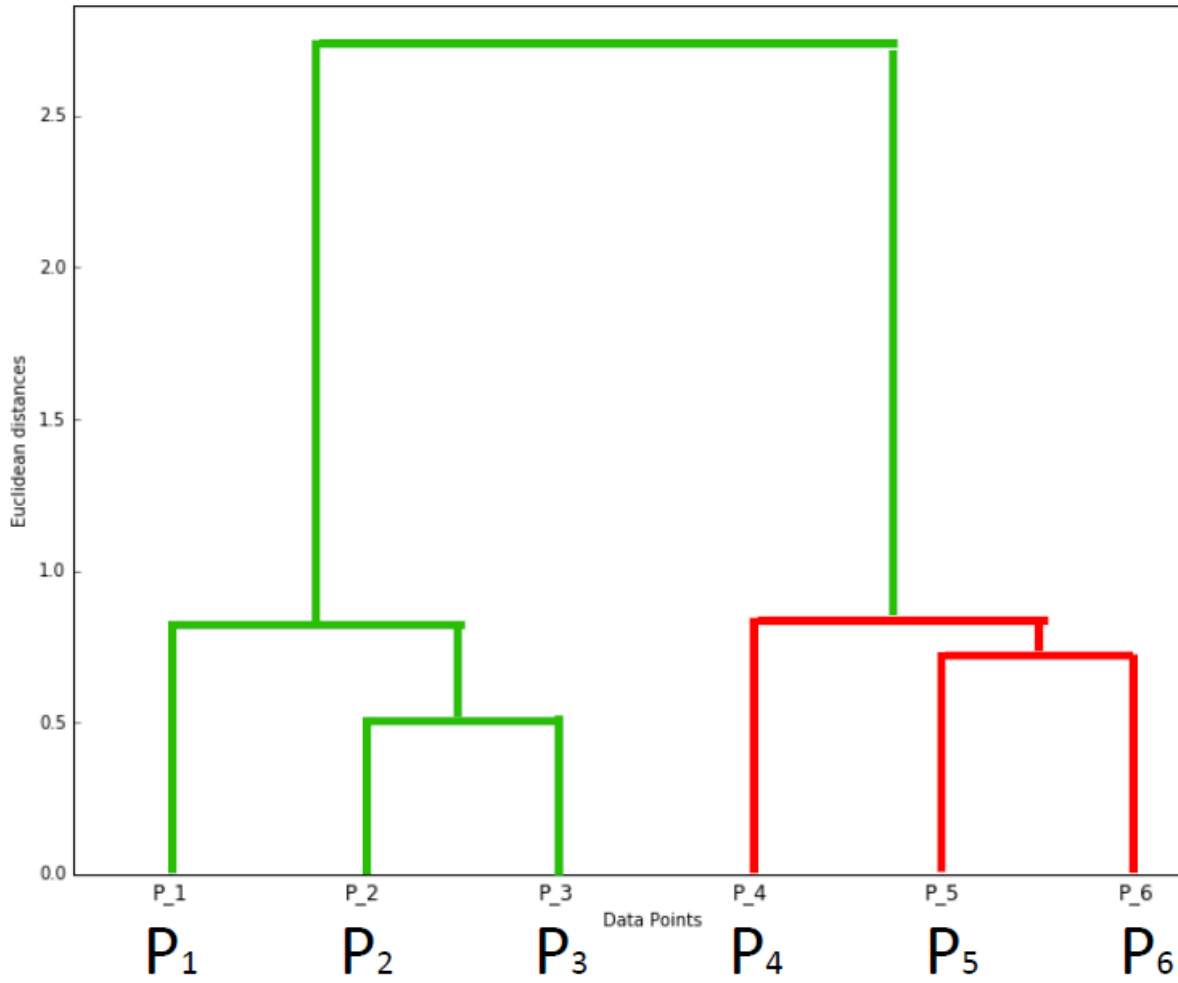
How Do Dendrograms works?



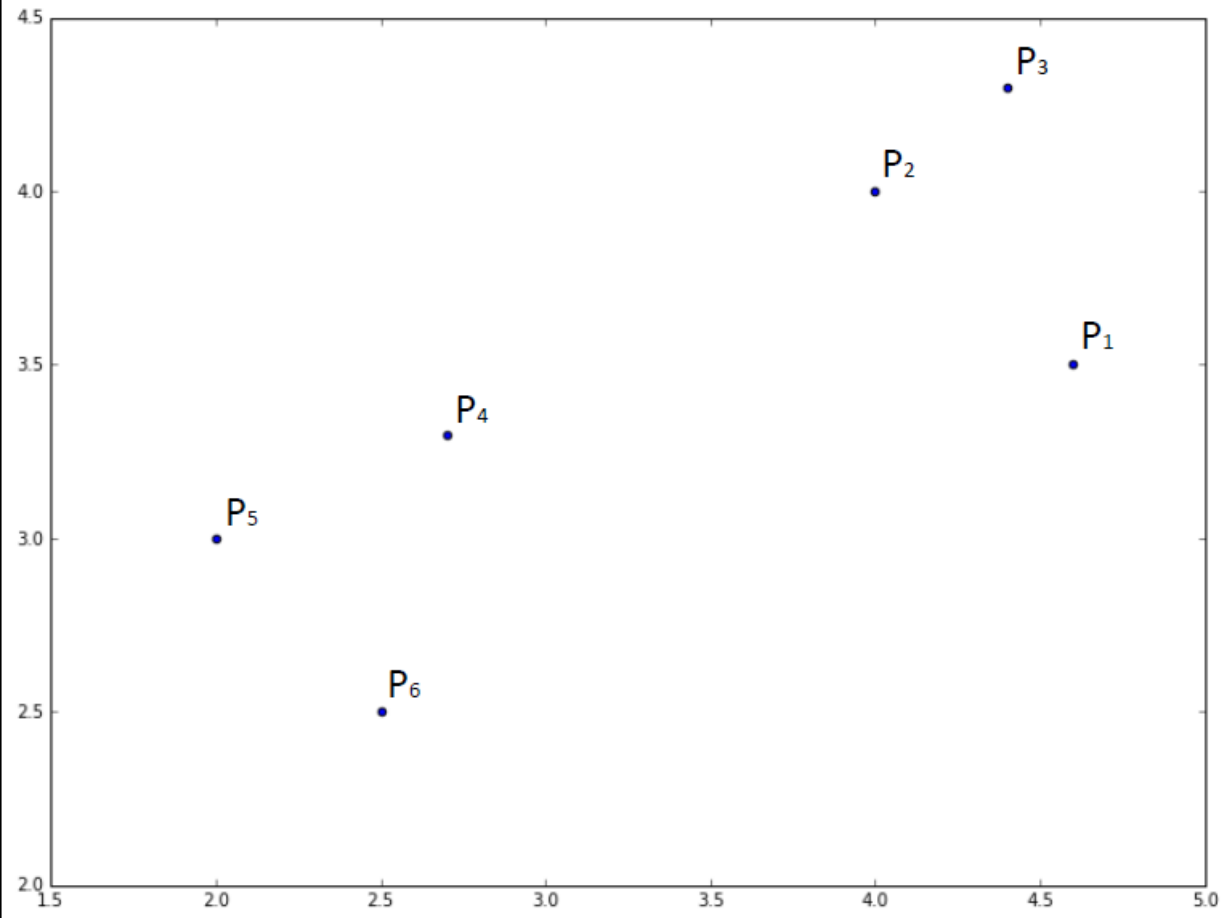


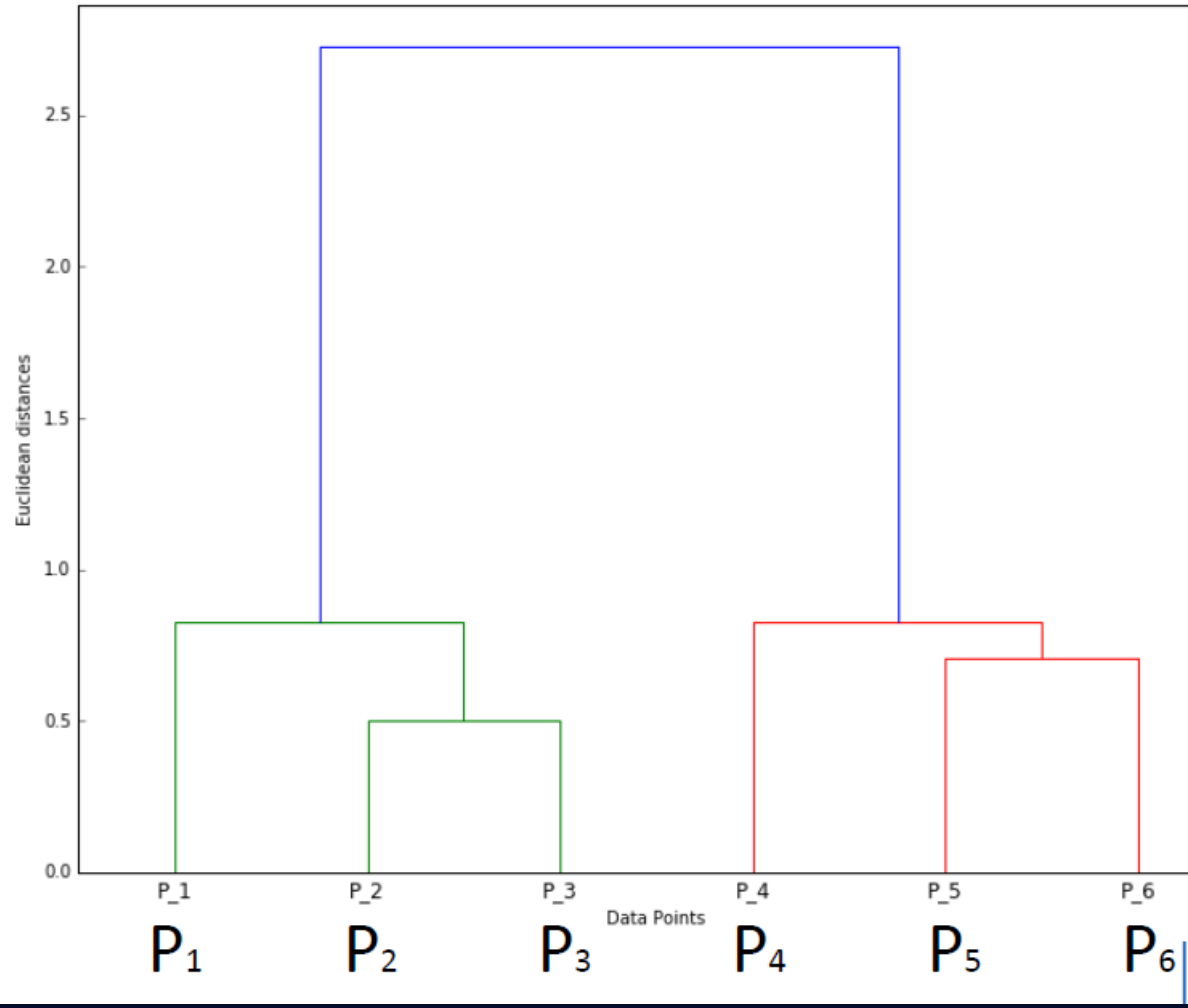
How Do Dendograms works?



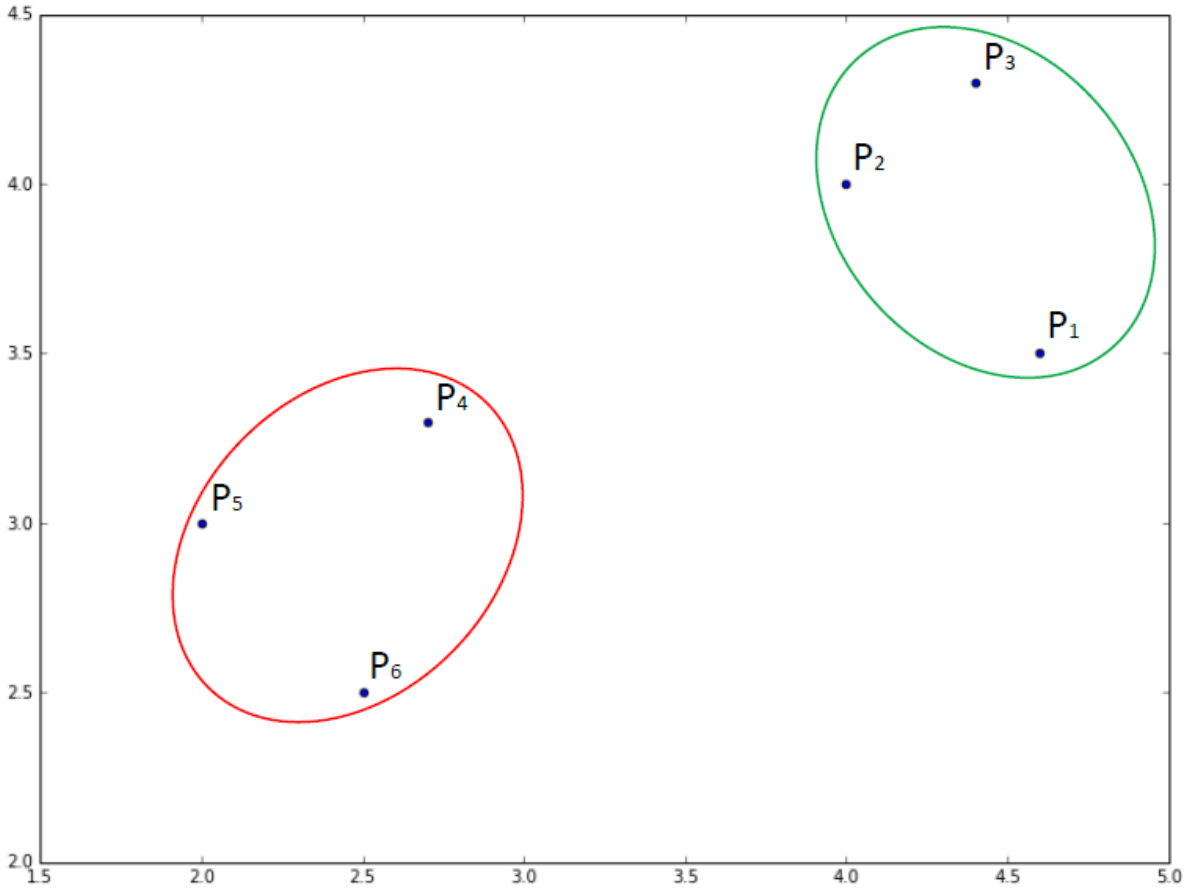


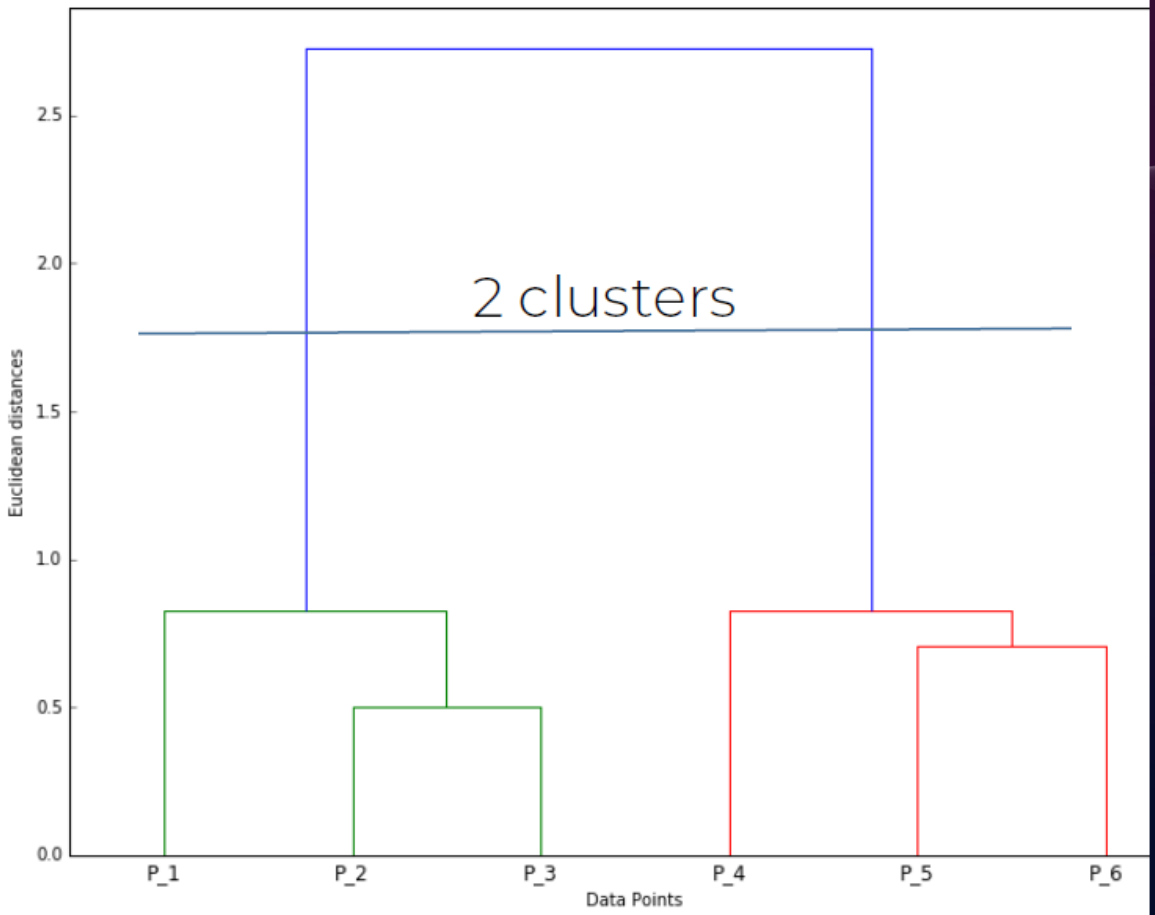
How Do Dendrograms works?



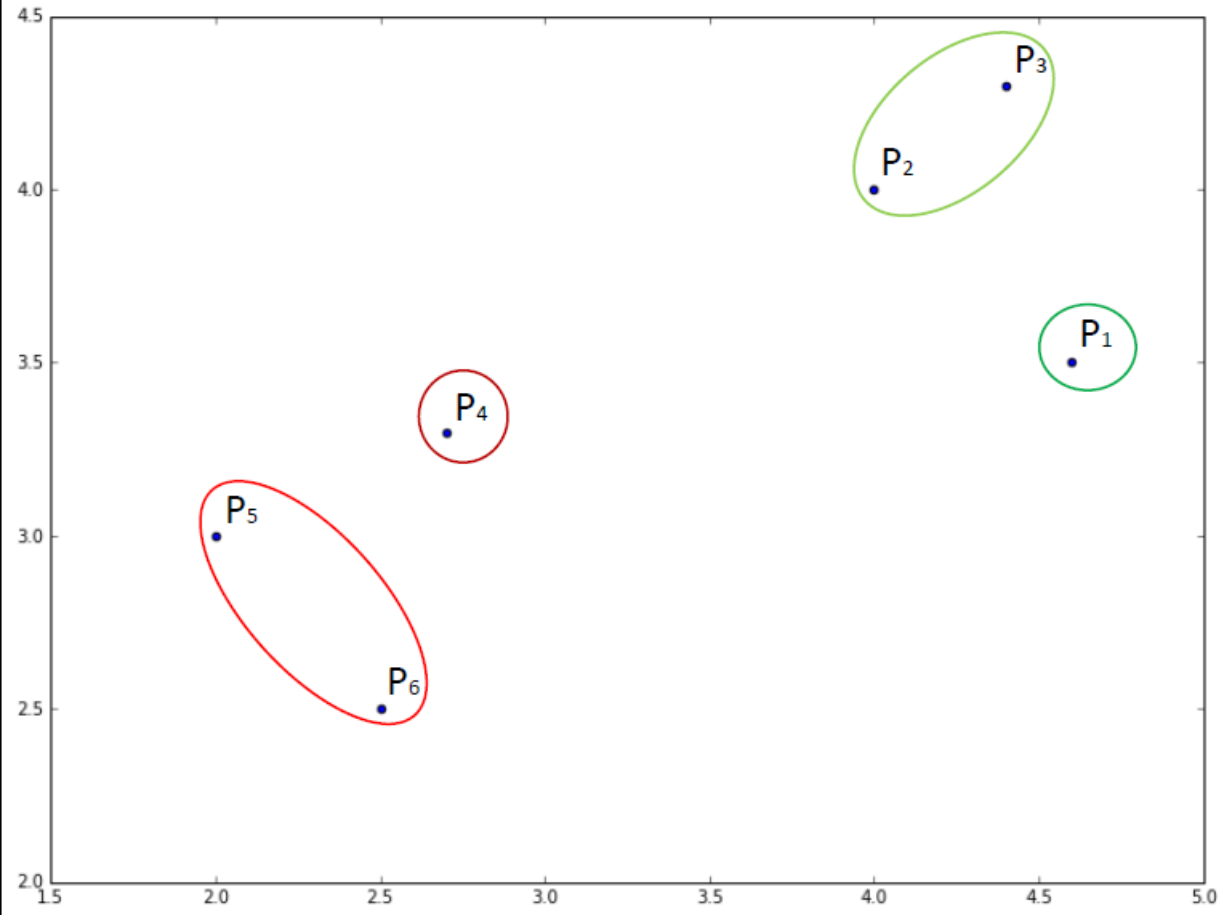


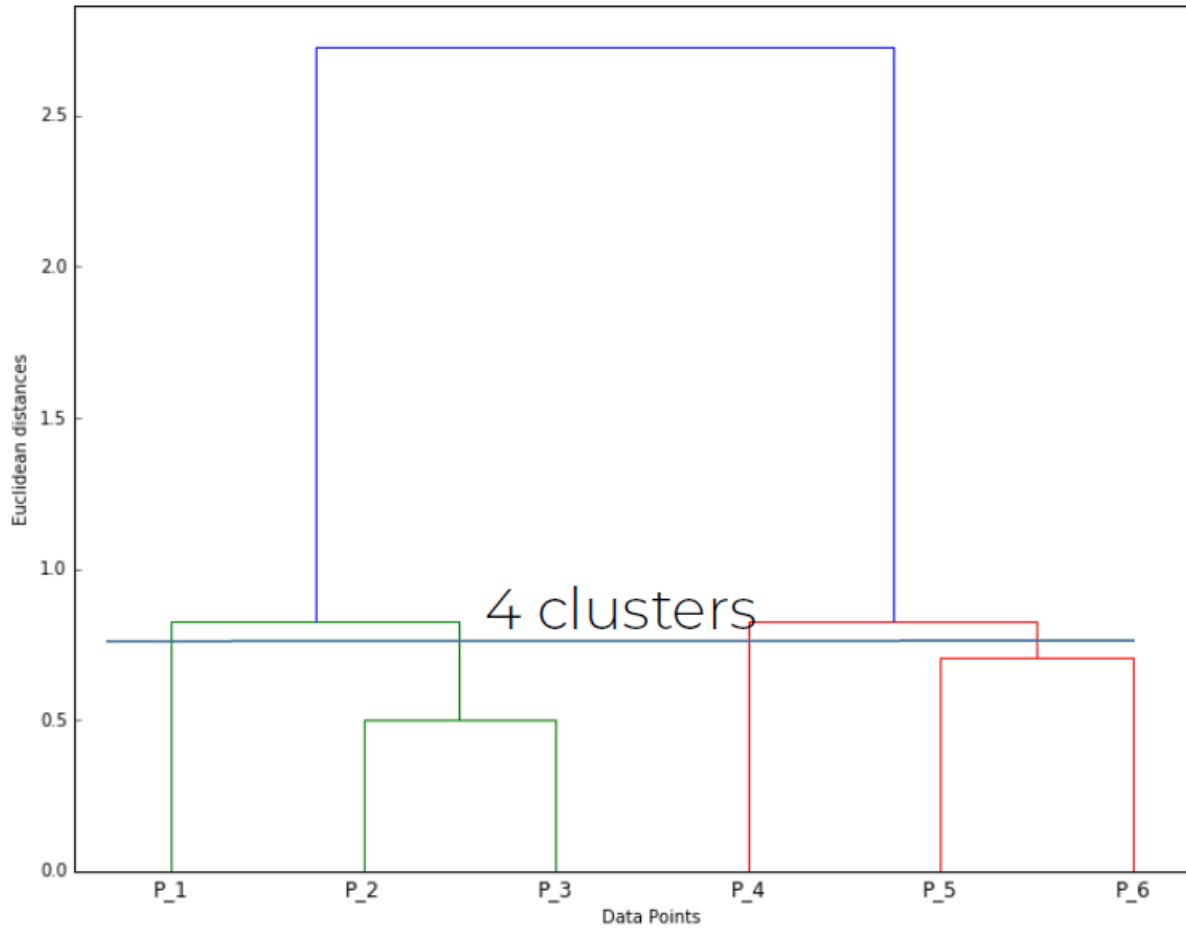
Dendrograms : Two Clusters



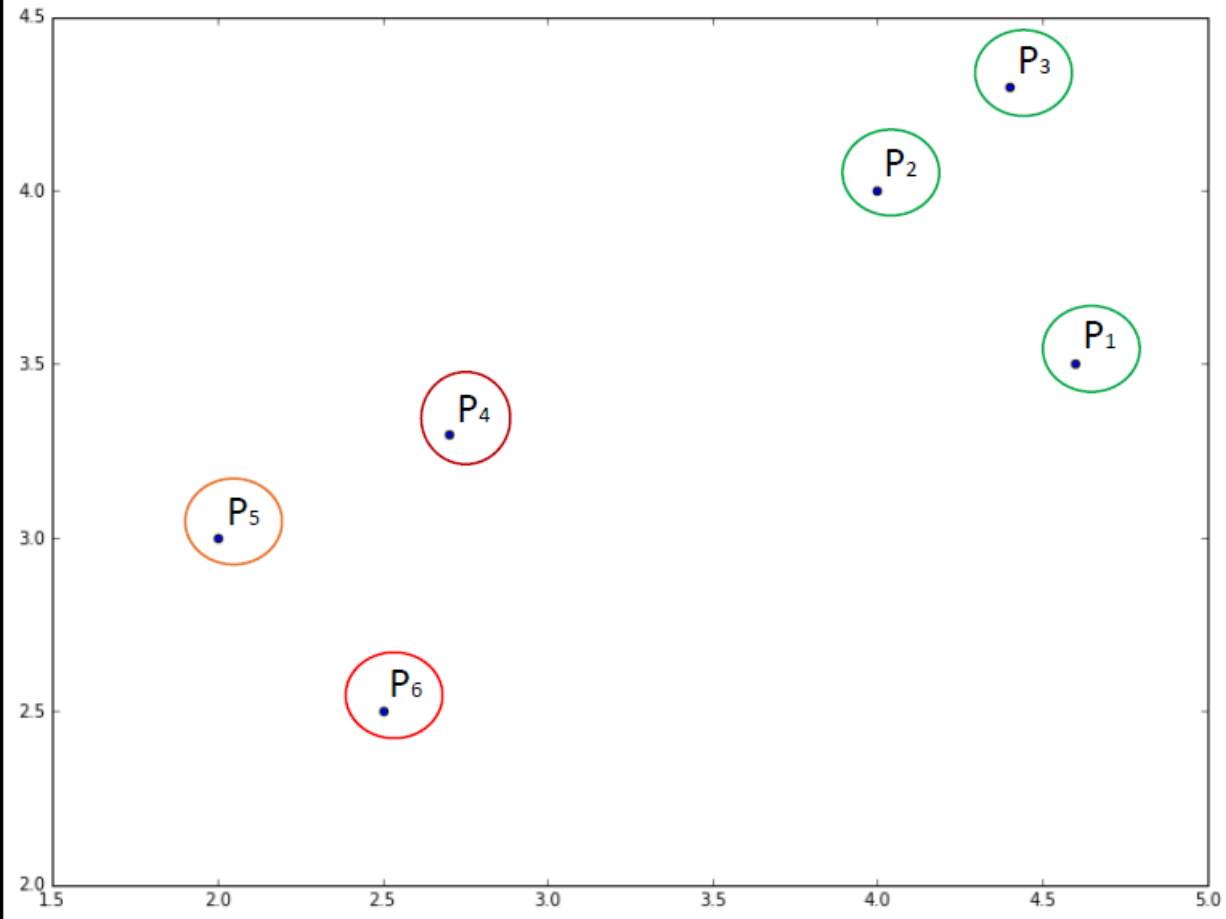


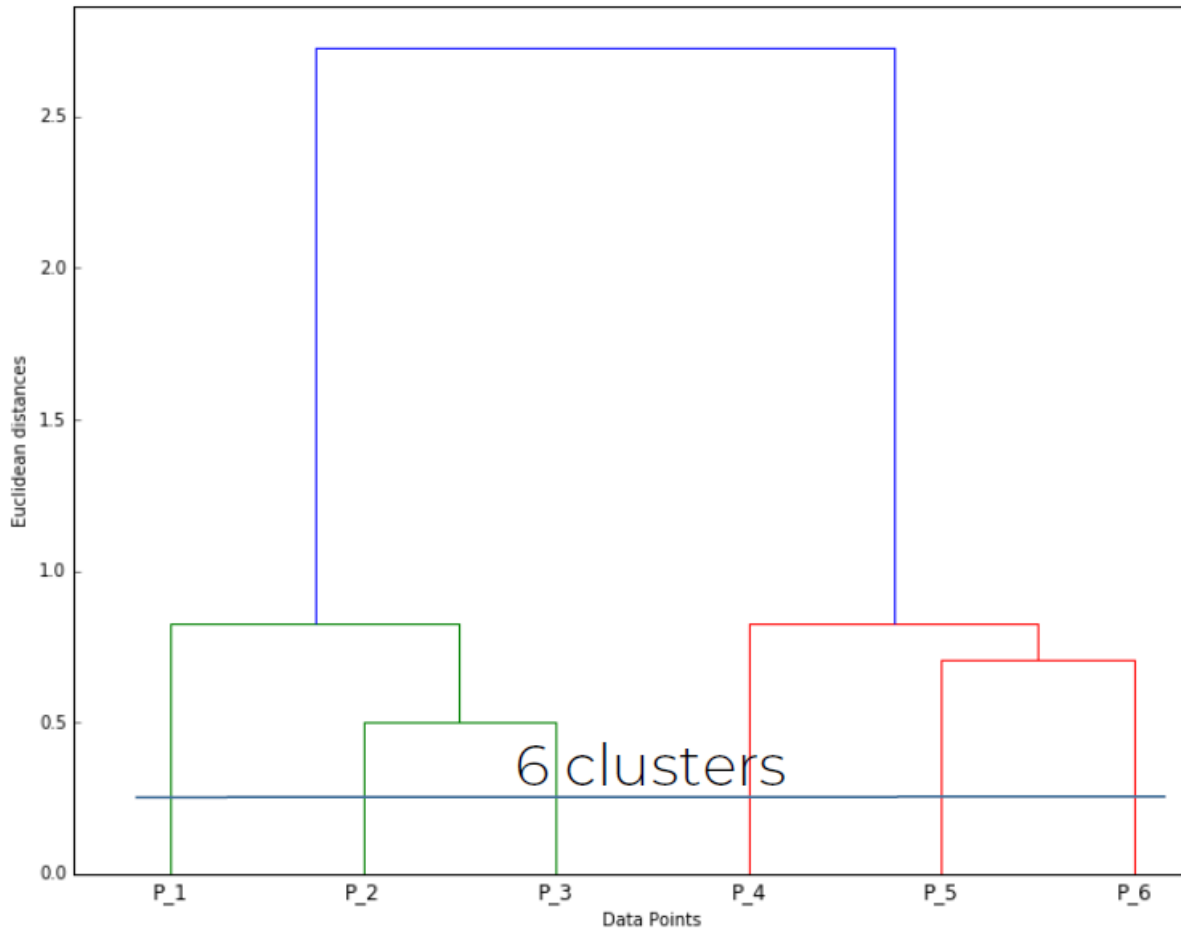
Dendrograms : Four Clusters



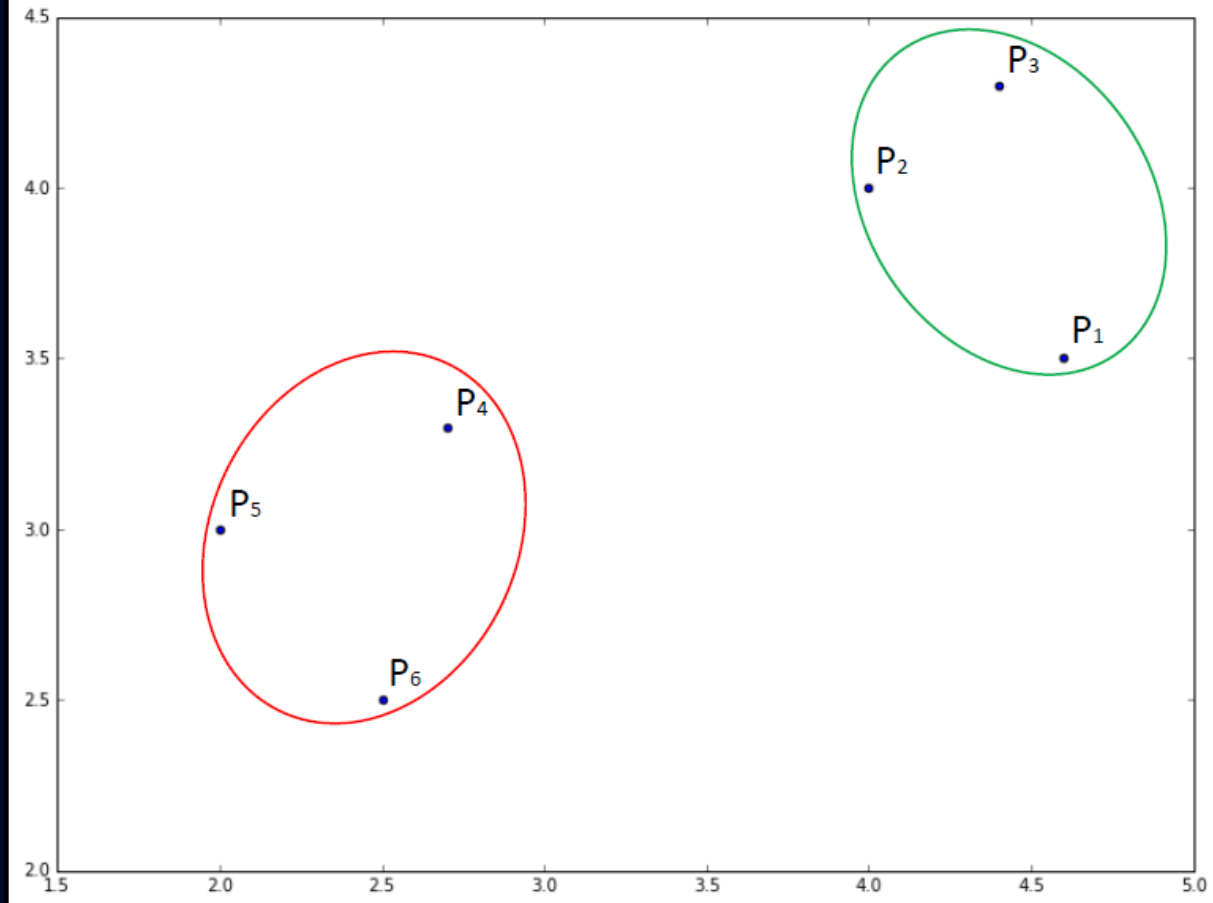


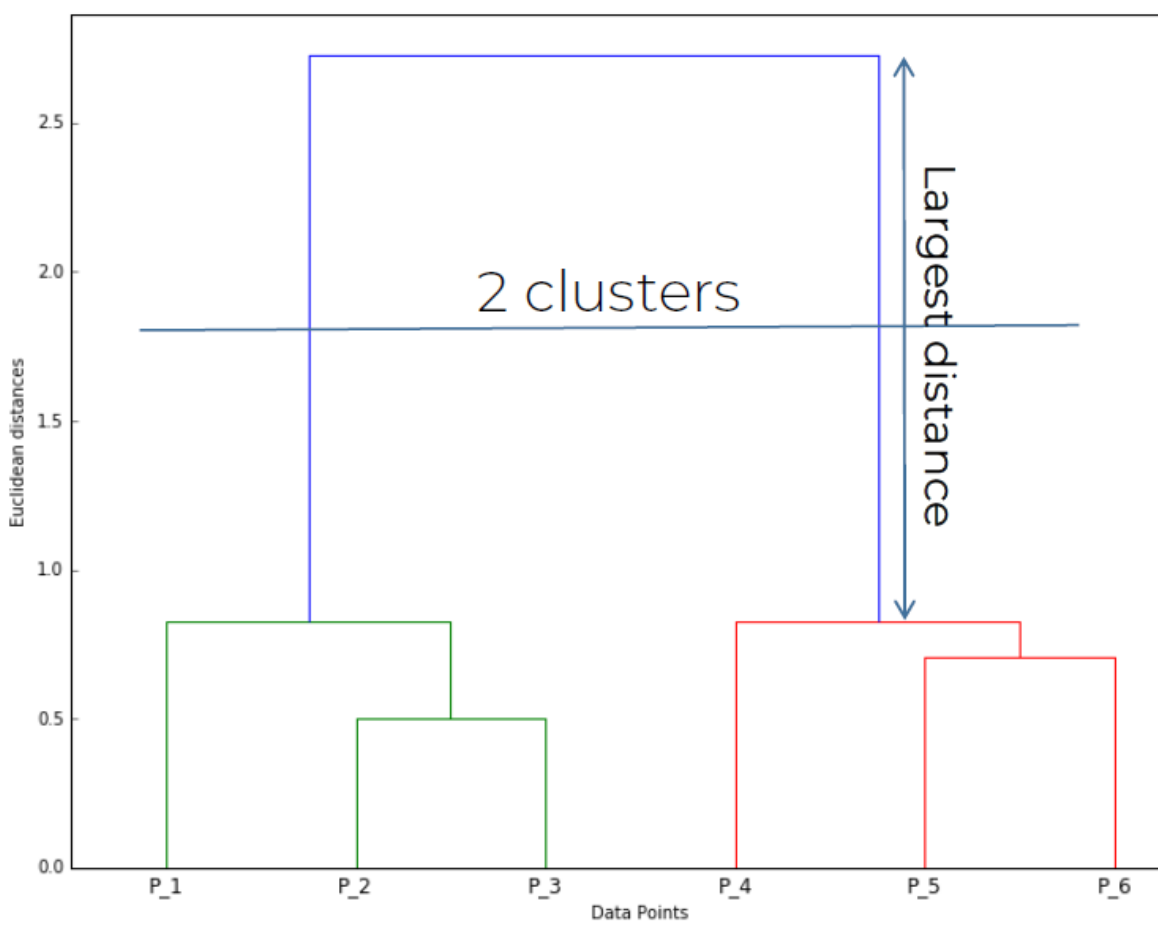
Dendrograms : Six Clusters





Dendrograms : Optimal #of Clusters





What is Deep Learning?

Artificial Intelligence

Any technique that enables computers to mimic human behavior

Machine Learning

Ability to learn without explicitly being programmed

Deep Learning

Extract patterns from data using neural networks

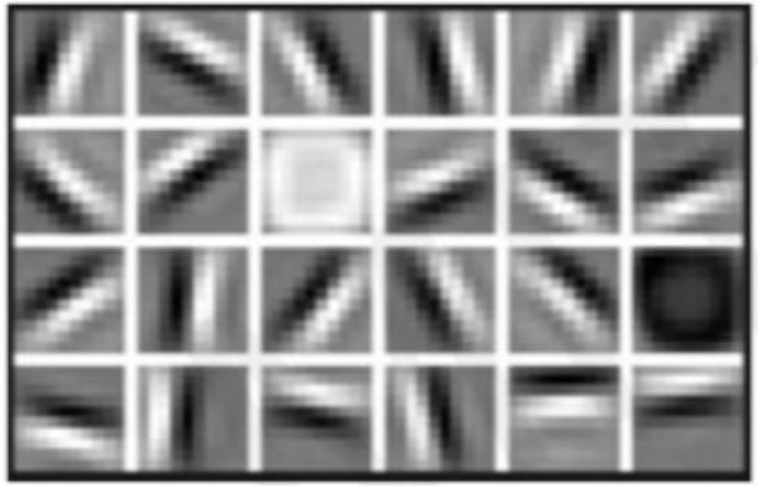
Teaching computers how to **learn a task** directly from **raw data**

Why Deep Learning?

Hand engineered features are time consuming, brittle, and not scalable in practice

Can we learn the **underlying features** directly from data?

Low Level Features



Lines & Edges

Mid Level Features

Eyes & Nose & Ears

High Level Features

Facial Structure


CERTIFIED
DISKETTE
ATARIWRITER
SELECT LETTER _
SONY
ATARI 800

1956

1980

2017

SanDisk Ultra 256GB MicroSDXC UHS-I Card with Adapter (SDSQUNI-256G-GN6MA).

by SanDisk

$149.99 ~~$199.99~~ Prime

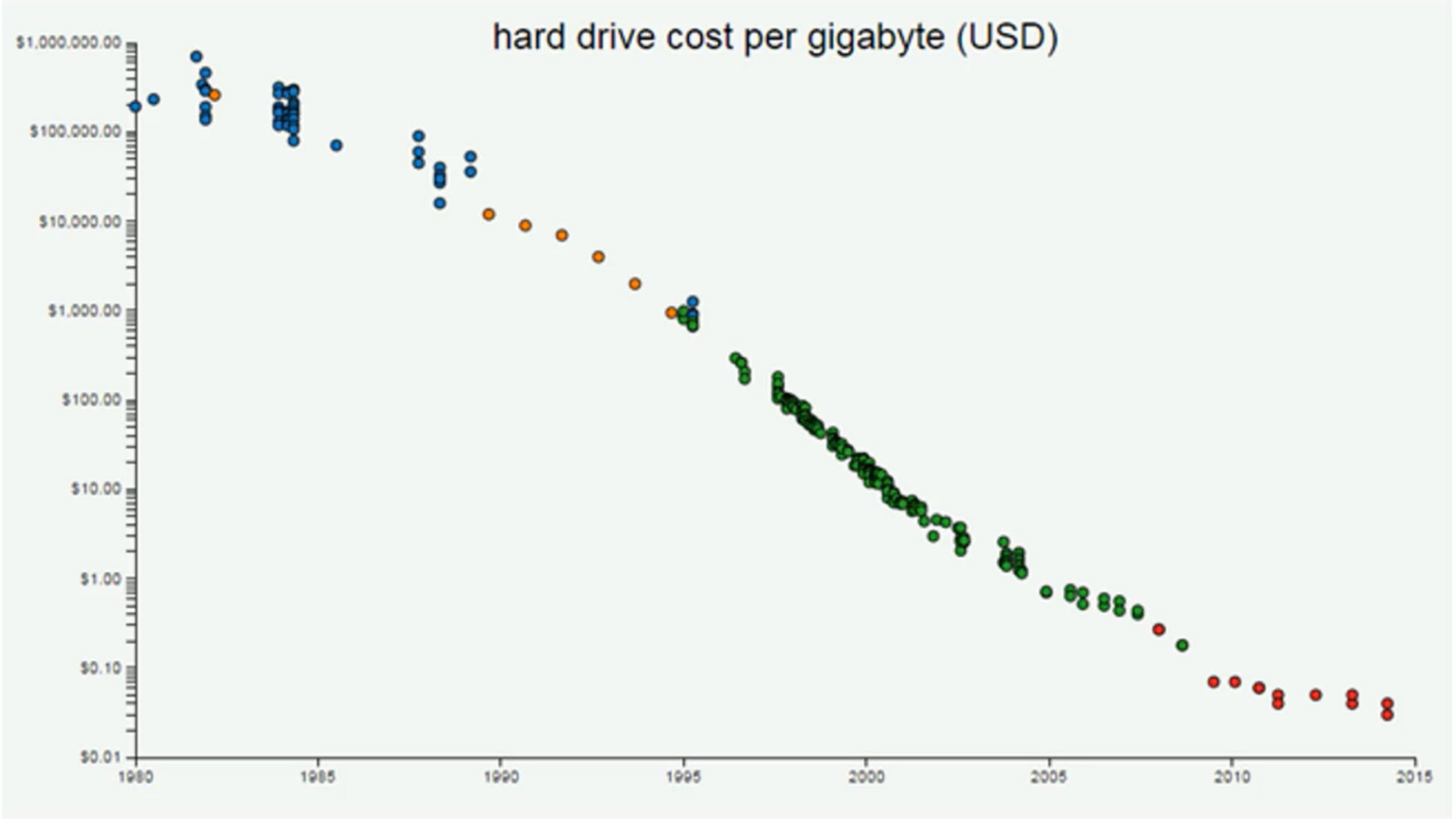
Get it by **Wednesday, Mar 22**
FREE Shipping on eligible orders

More Buying Choices
$147.99 (10 new offers)

4.5 stars 21,224

Product Features
256GB capacity breakthrough

Log-scale


hard drive cost per gigabyte (USD)
$1,000,000.00
$100,000.00
$10,000.00
$1,000.00
$100.00
$10.00
$1.00
$0.10
$0.01
1980
1985
1990
1995
2000
2005
2010
2015

Why Now?

Neural Networks date back decades, so why the dominance?

1952 Stochastic Gradient Descent

1958 Perceptron
- Learnable Weights

1986 Backpropagation
- Multi-Layer Perceptron

1995 Deep Convolutional NN
- Digit Recognition

1. Big Data
- Larger Datasets
- Easier Collection & Storage

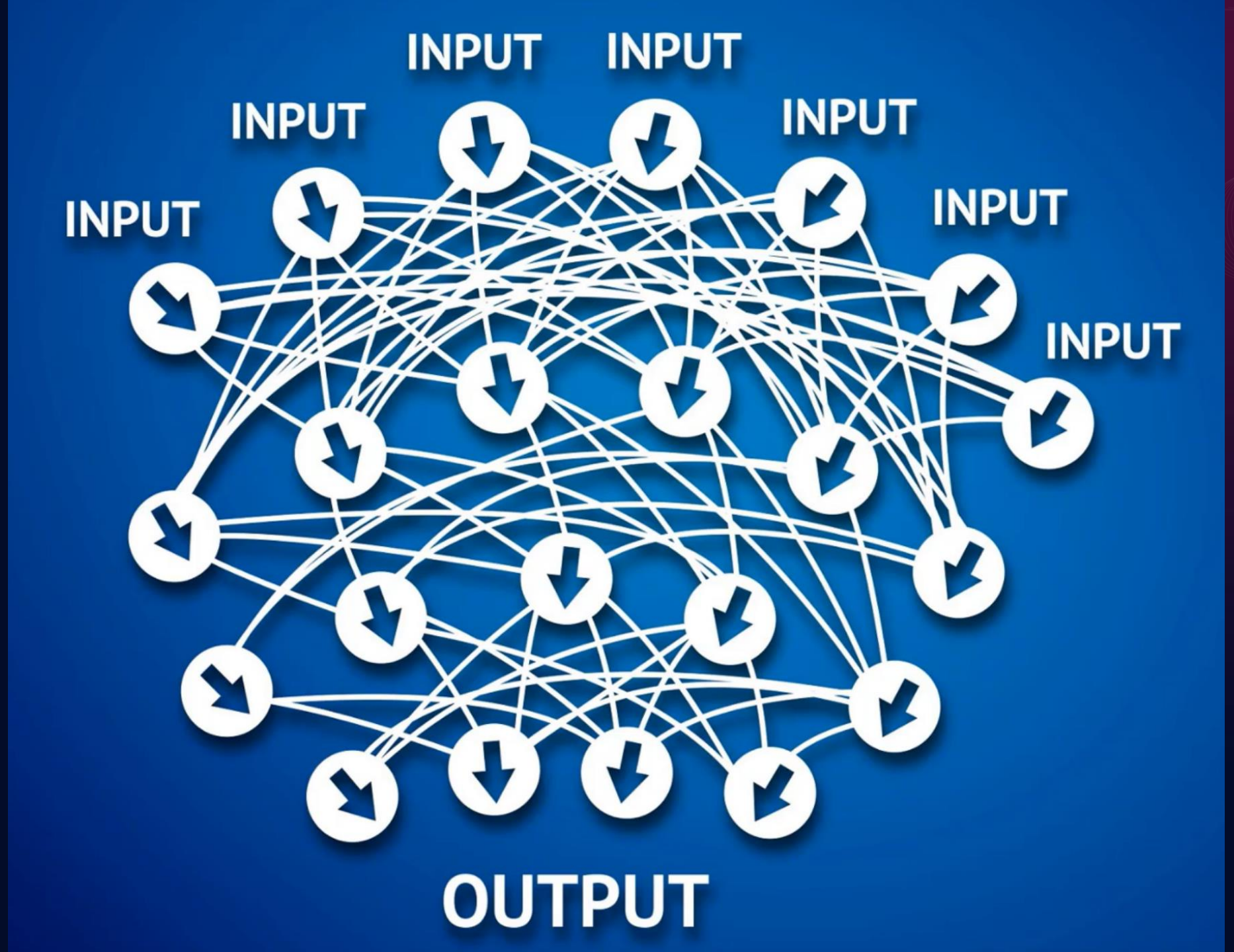
IMAGENET

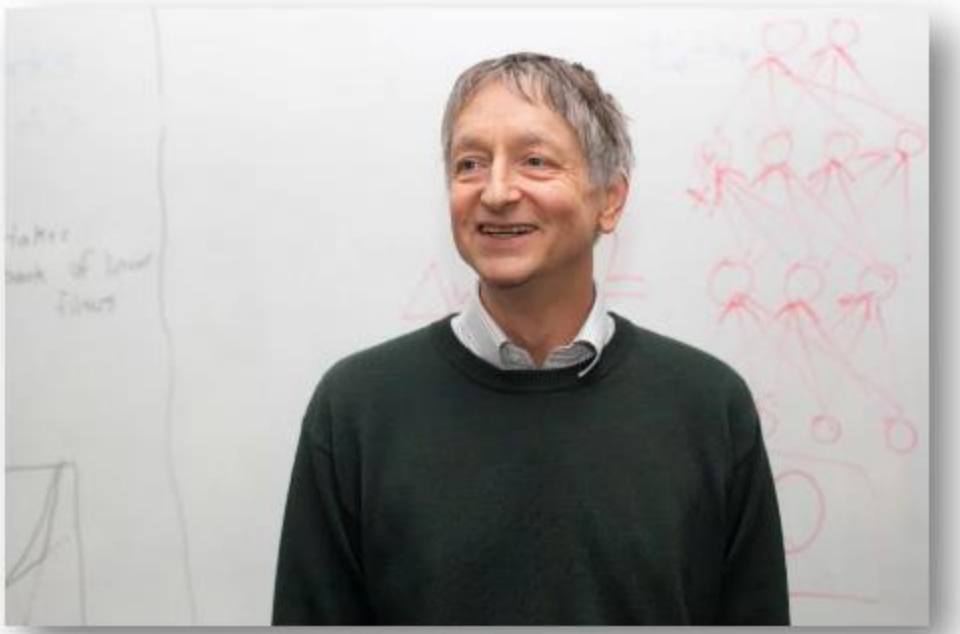
WIKIPEDIA

2. Hardware
- Graphics Processing Units (GPUs)
- Massively Parallelizable

3. Software
- Improved Techniques
- New Models
- Toolboxes

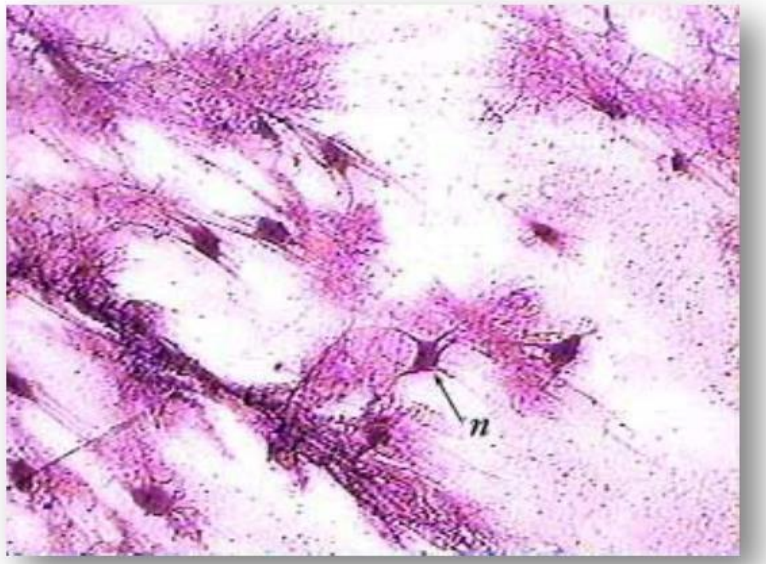
TensorFlow

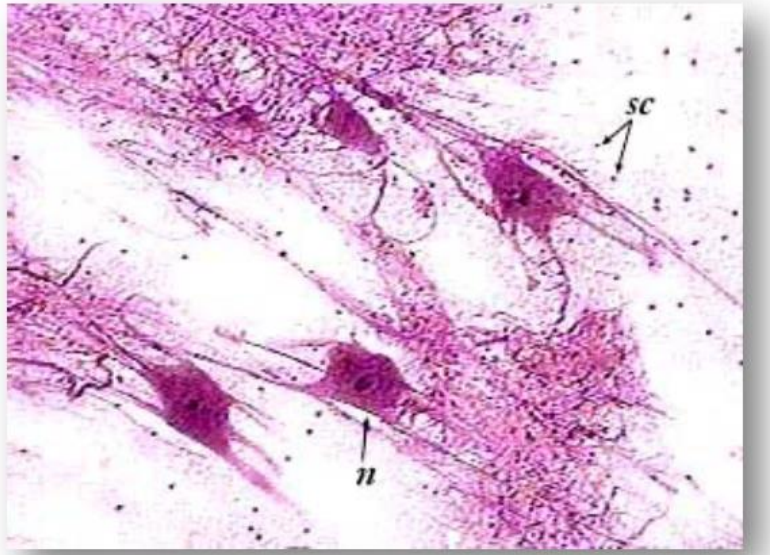

INPUT
INPUT
INPUT
INPUT
INPUT
INPUT
INPUT
OUTPUT



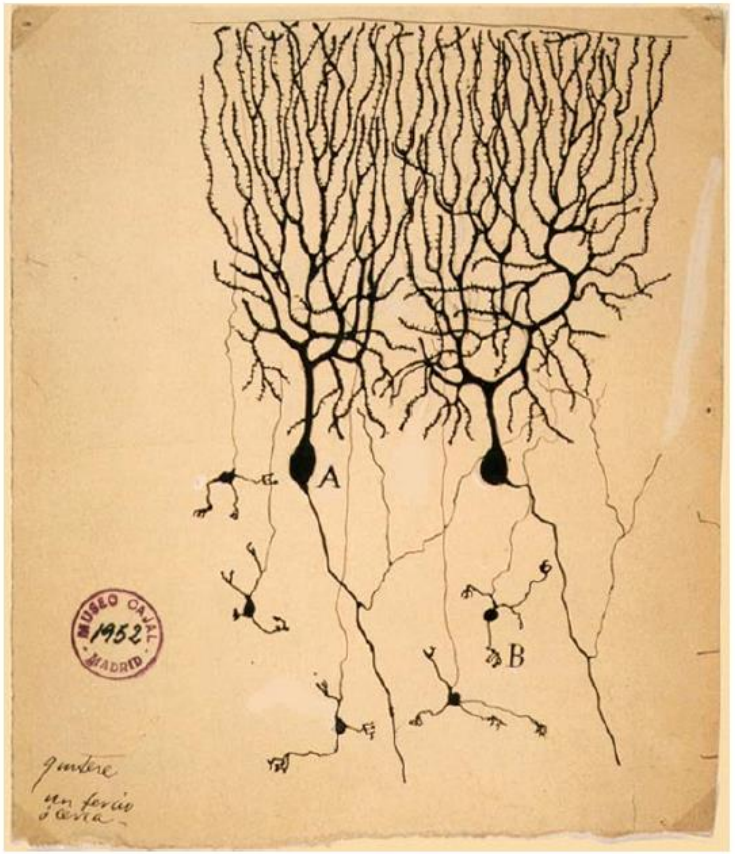
Geoffrey Hinton

The Neuron

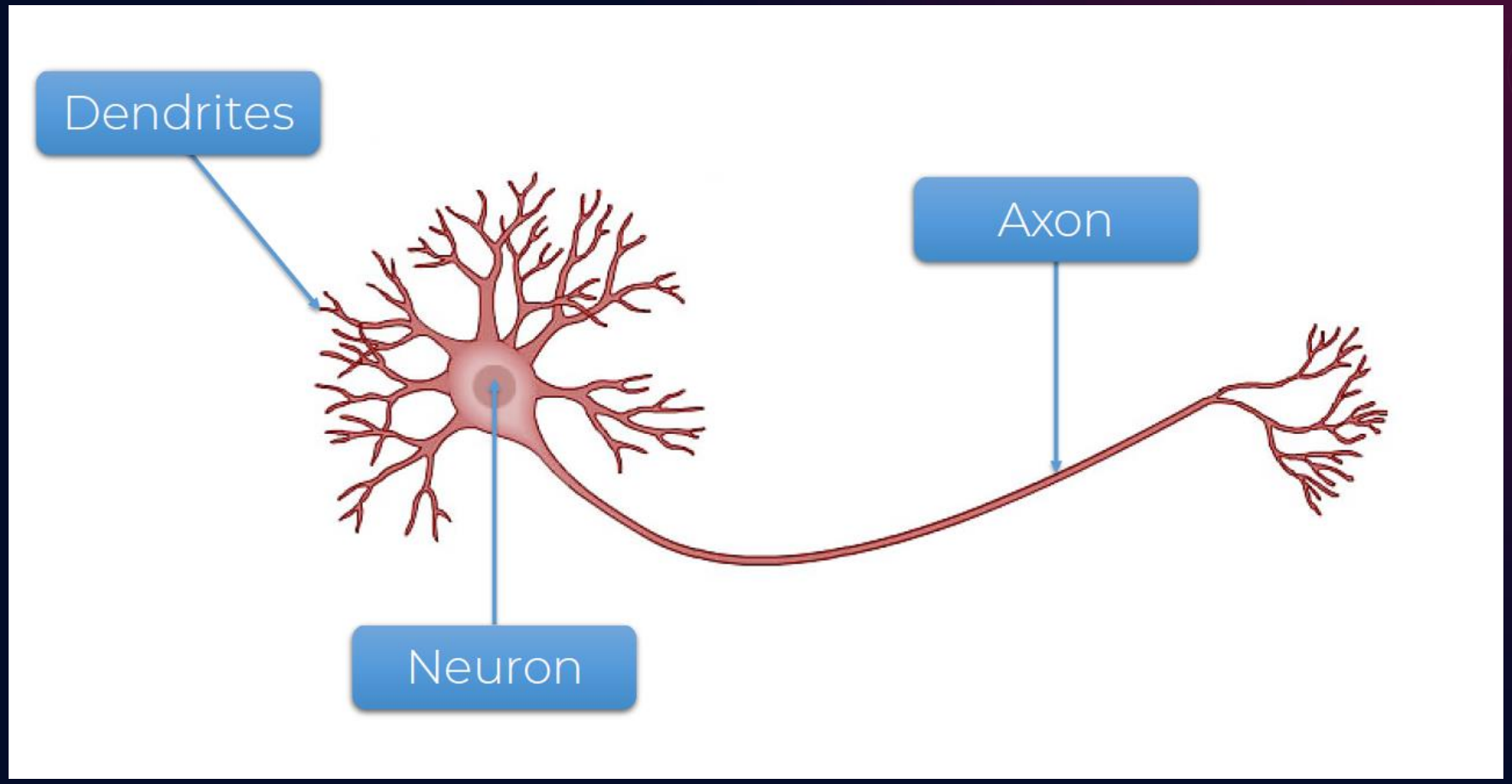




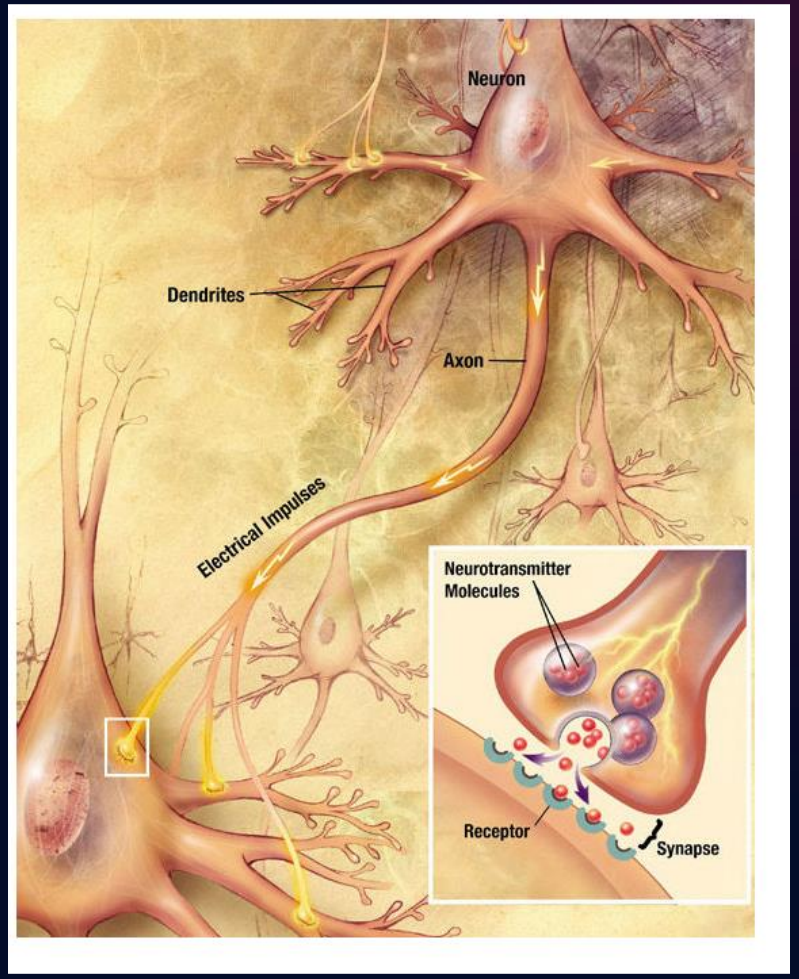
The Neuron



The Neuron



The Neuron


Neuron
Dendrites
Axon
Electrical Impulses
Neurotransmitter Molecules
Receptor
Synapse