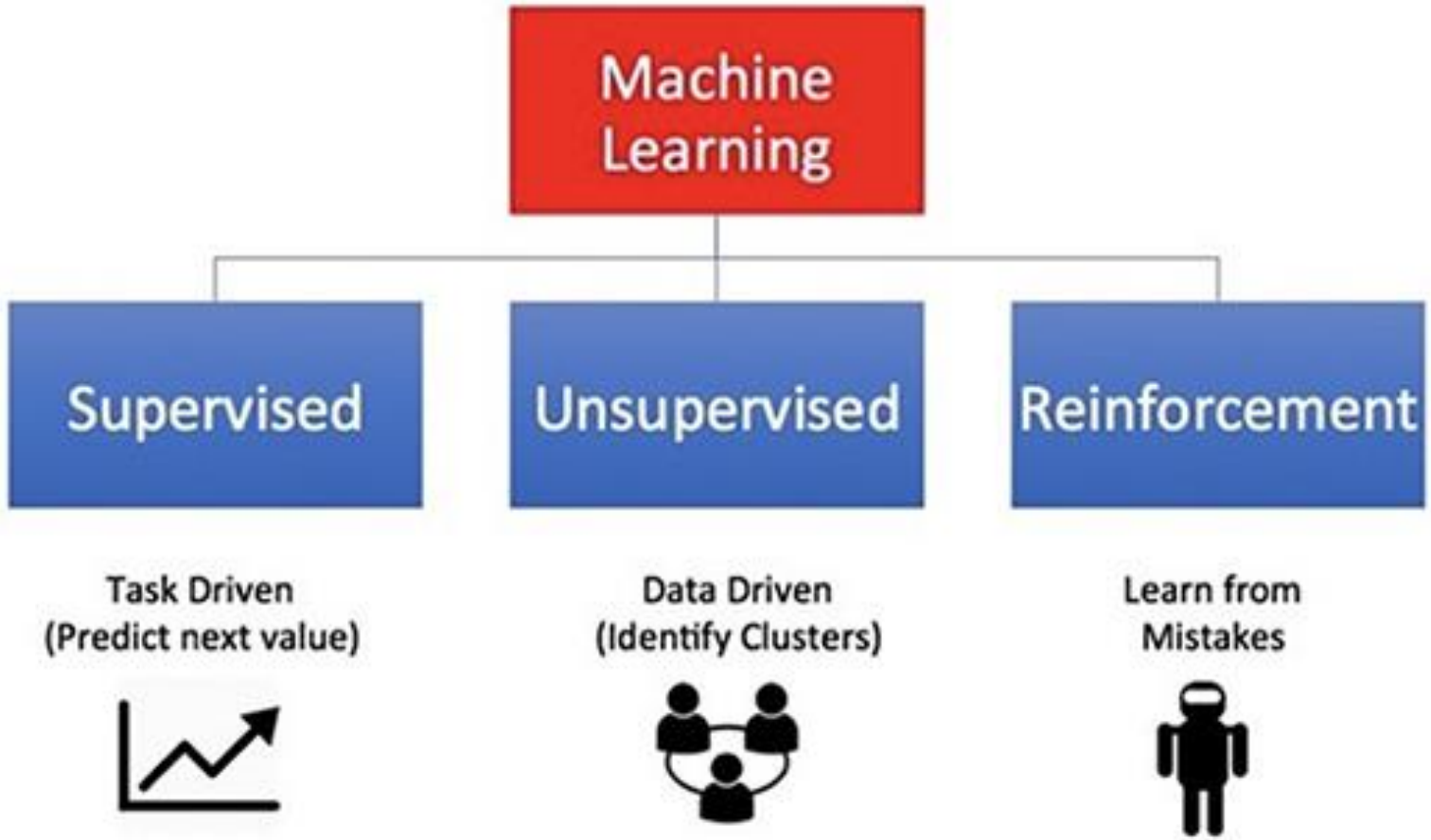
Machine Learning

- Unsupervised Learning -

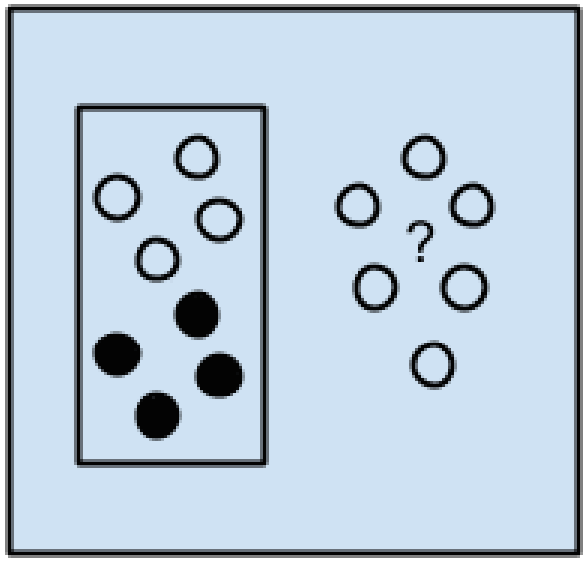
2022. 4. 29

정 준 수 PhD

Types of Machine Learning

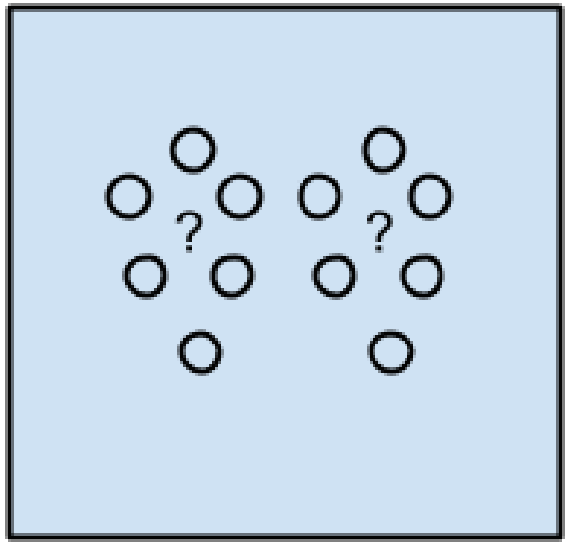


Supervised Learning



Example problems are classification and regression.

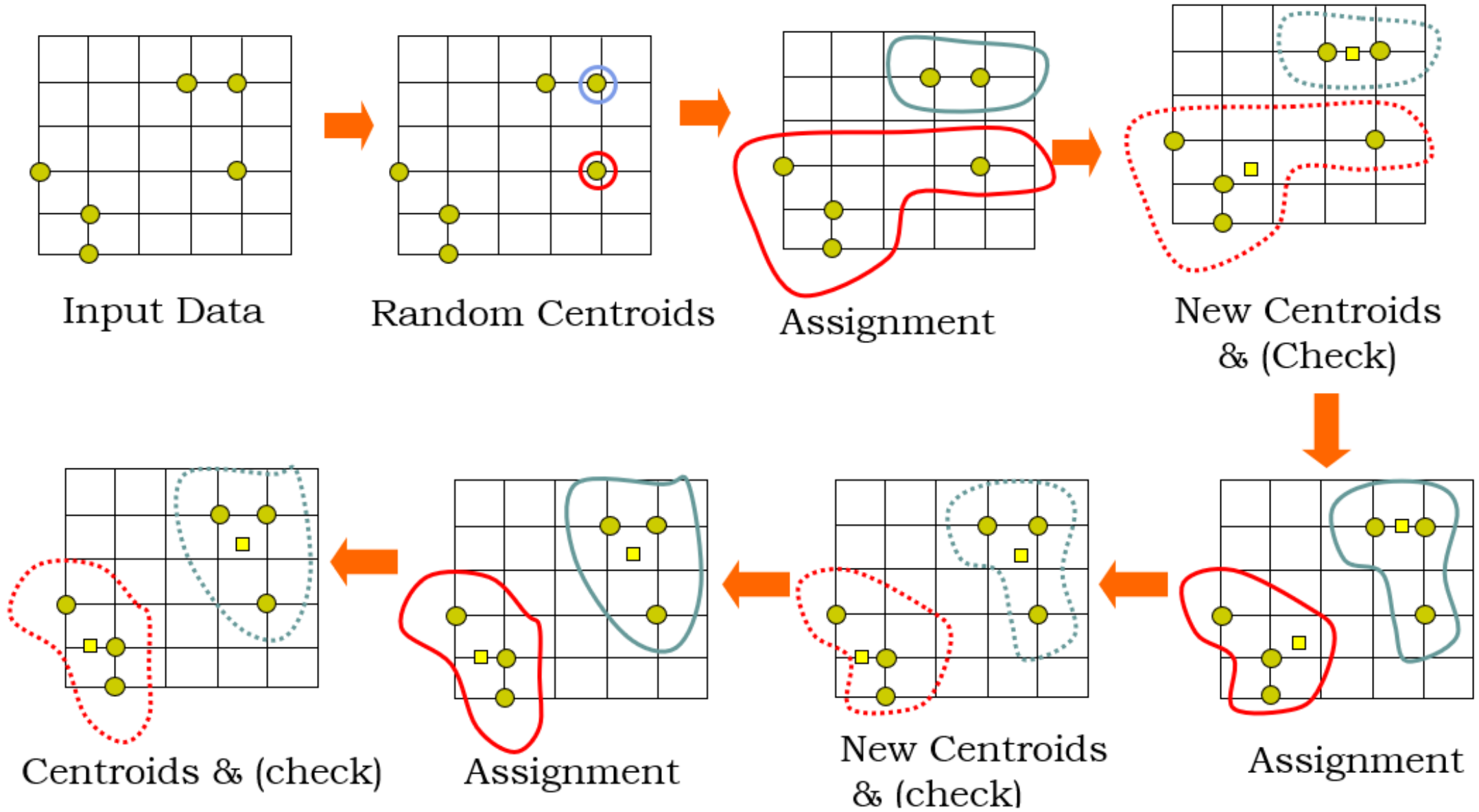
Unsupervised Learning



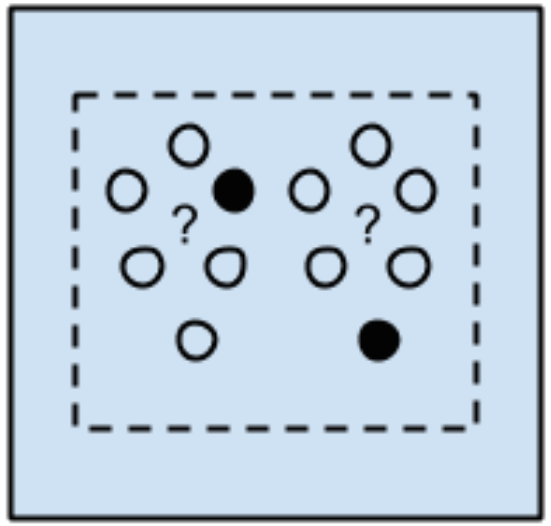
Unsupervised Learning Algorithms

Input data is not labeled and does not have a known result.

Example problems are clustering, dimensionality reduction and association rule learning.


Input Data
Random Centroids
Assignment
New Centroids
& (Check)
Assignment
New Centroids
& (check)
Assignment
Centroids & (check)

Semi-Supervised Learning



Semi-supervised
Learning Algorithms

Input data is a mixture of labeled and unlabelled examples.

Example problems are classification and regression.

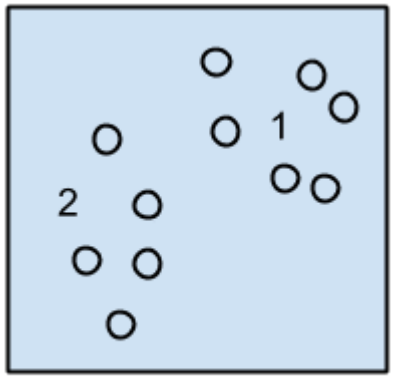
Clustering Algorithms

Clustering, like regression, describes the class of problem and the class of methods

Clustering methods are typically organized by the modeling approaches such as centroid-based and hierarchal. All methods are concerned with using the inherent structures in the data to best organize the data into groups of maximum commonality.

The most popular clustering algorithms are:

k-Means
k-Medians
Expectation Maximisation (EM)
Hierarchical Clustering



Clustering Algorithms

[출처] https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

K-means

https://scikit-learn.org/stable/modules/clustering.html#k-means

In Depth: k-Means Clustering

https://jakevdp.github.io/PythonDataScienceHandbook/05.11-k-means.html

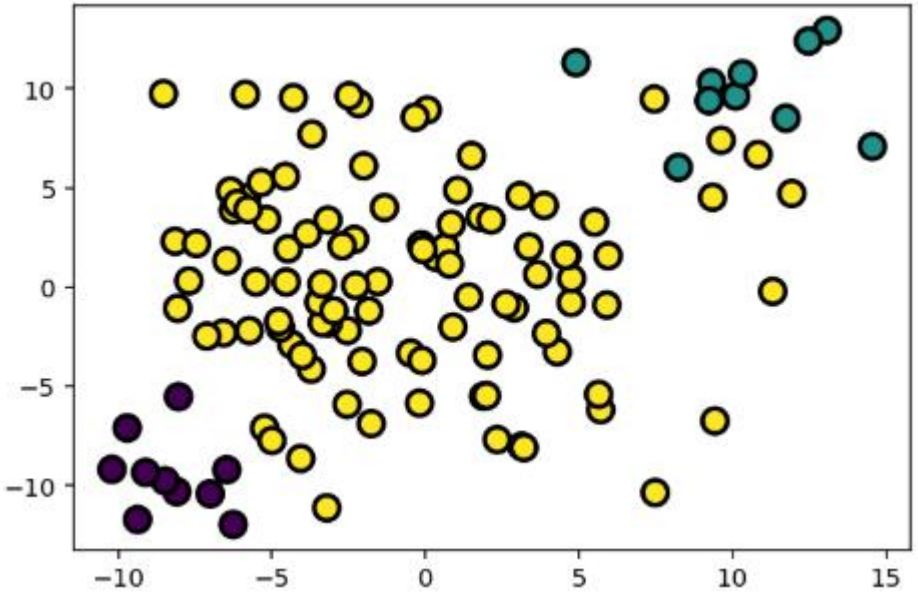
k-Means 클러스터링의 한계점

군의 특성이 다를 경우

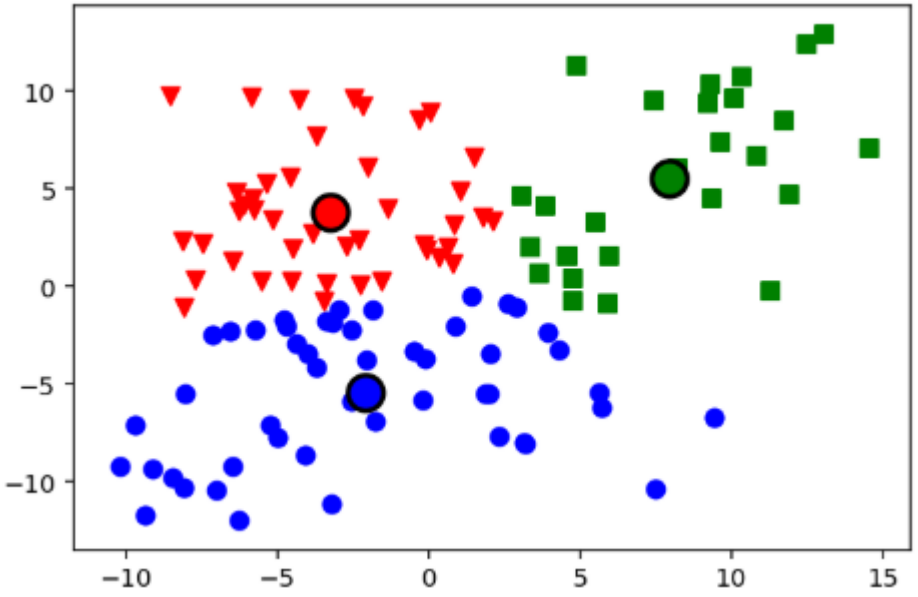
- 크기(Sizes)
- 밀도(Densities)
- 비 구형(Non-globular shapes)

이상치를 포함할 경우

군의 크기가 다를 경우

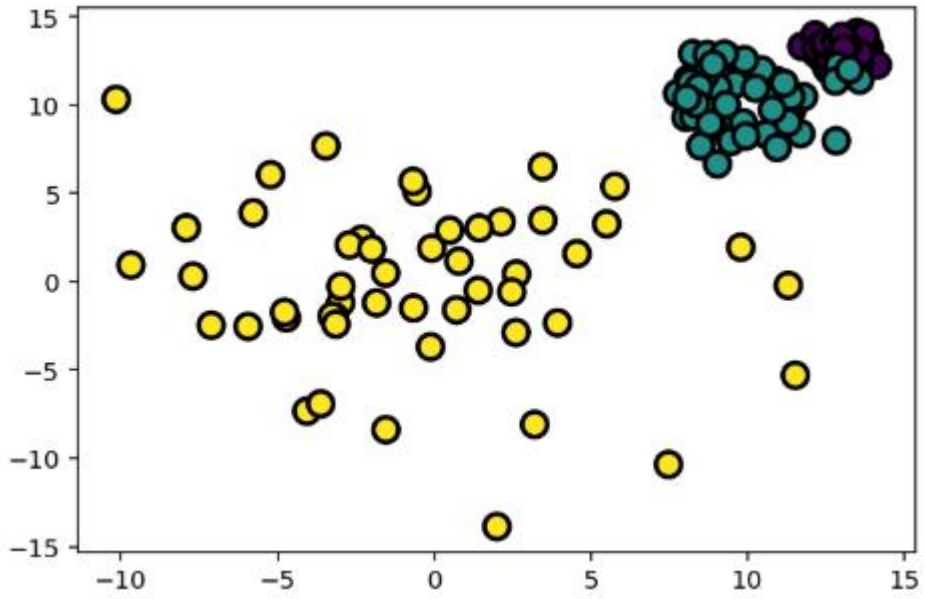


Original Points

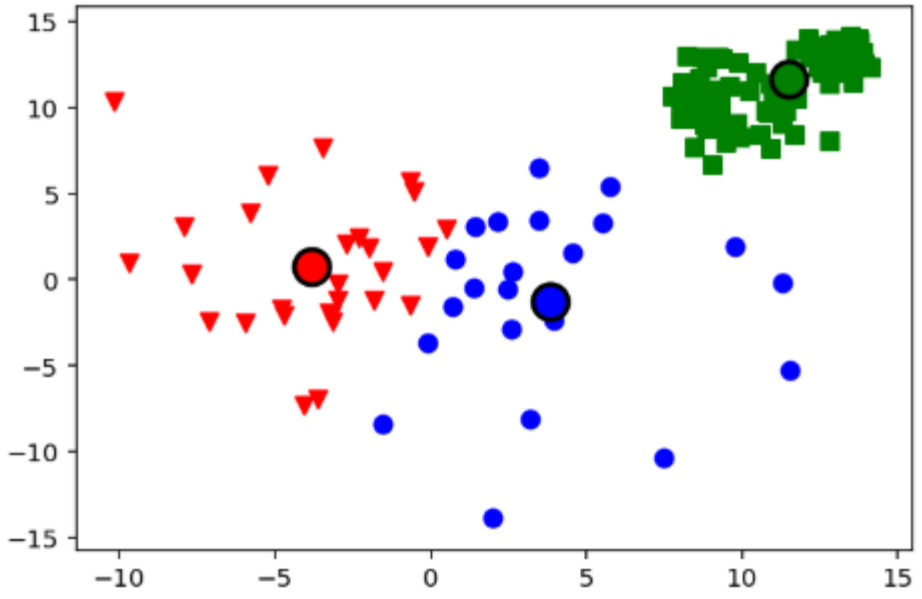


k-means (3 Clusters)

군의 밀도가 다를 경우

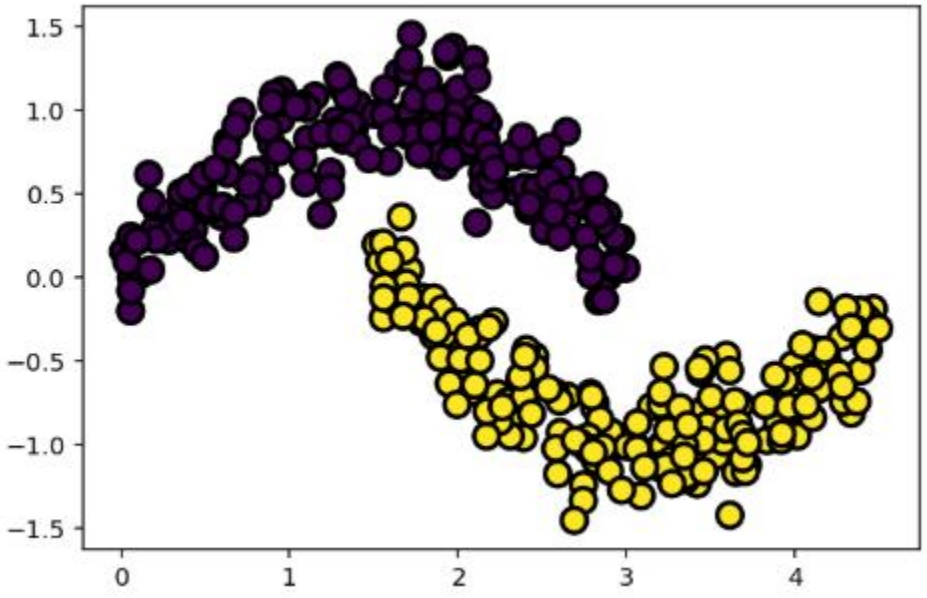


Original Points

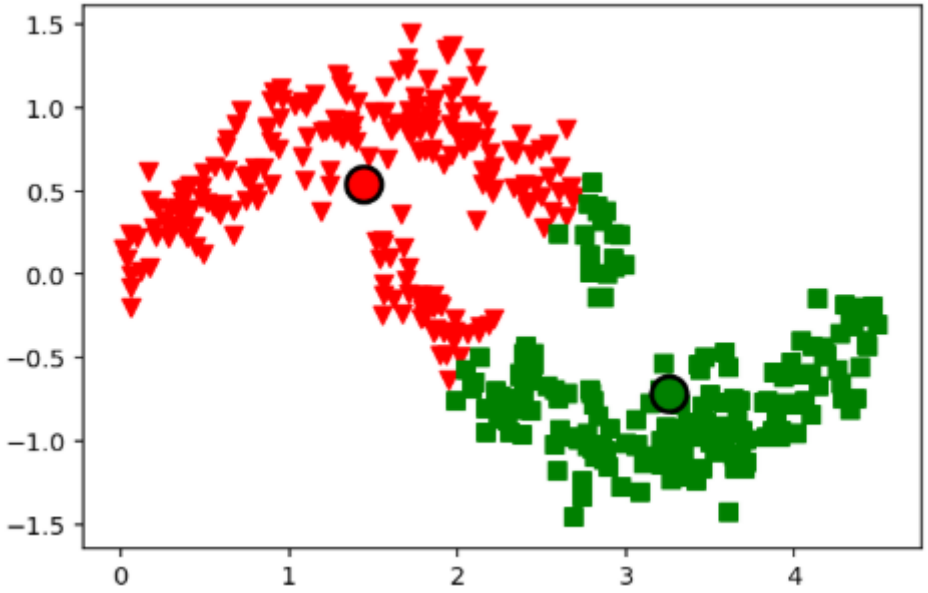


k-means (3 Clusters)

군이 구형이 아닌 경우

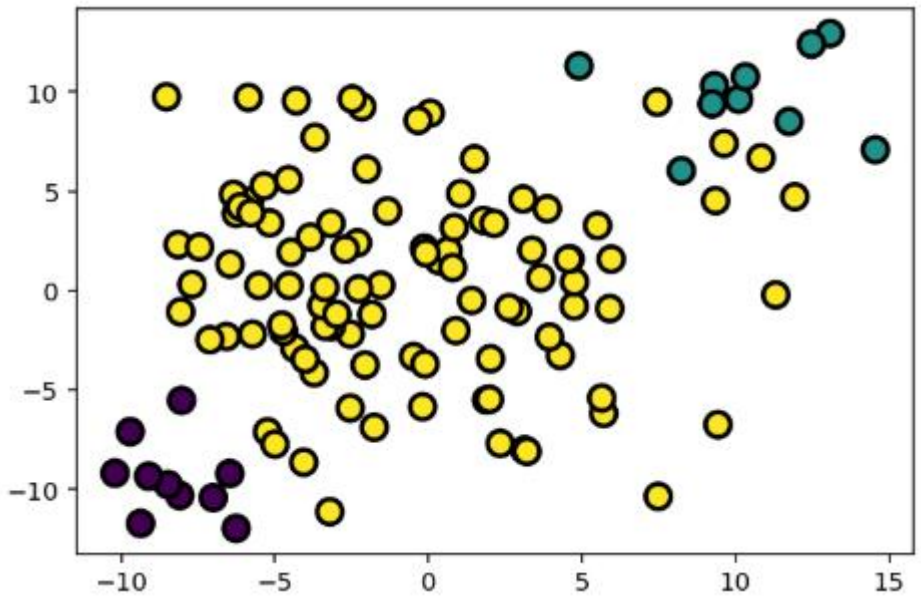


Original Points

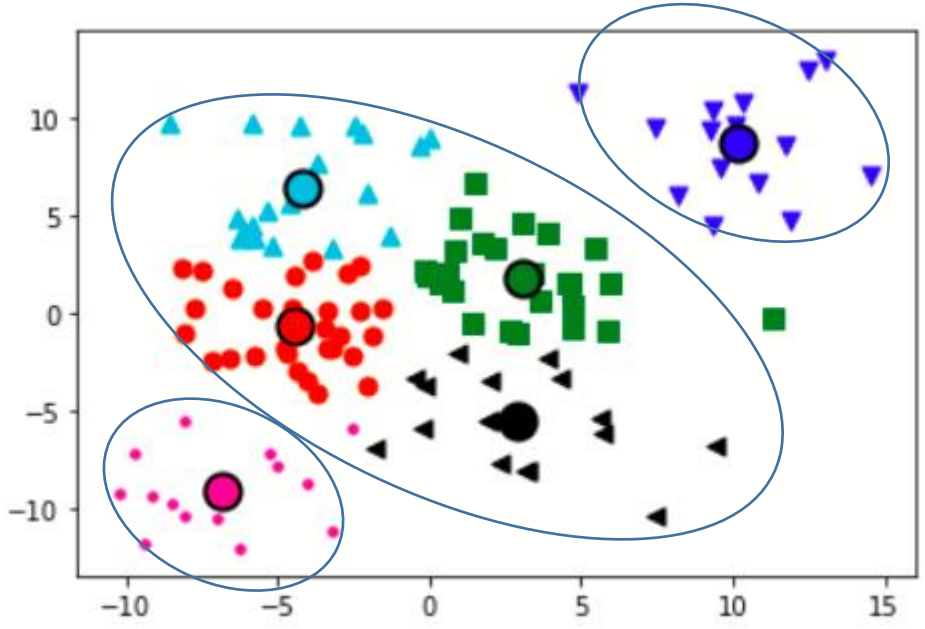


k-means (2 Clusters)

군의 크기가 다를 경우 해결

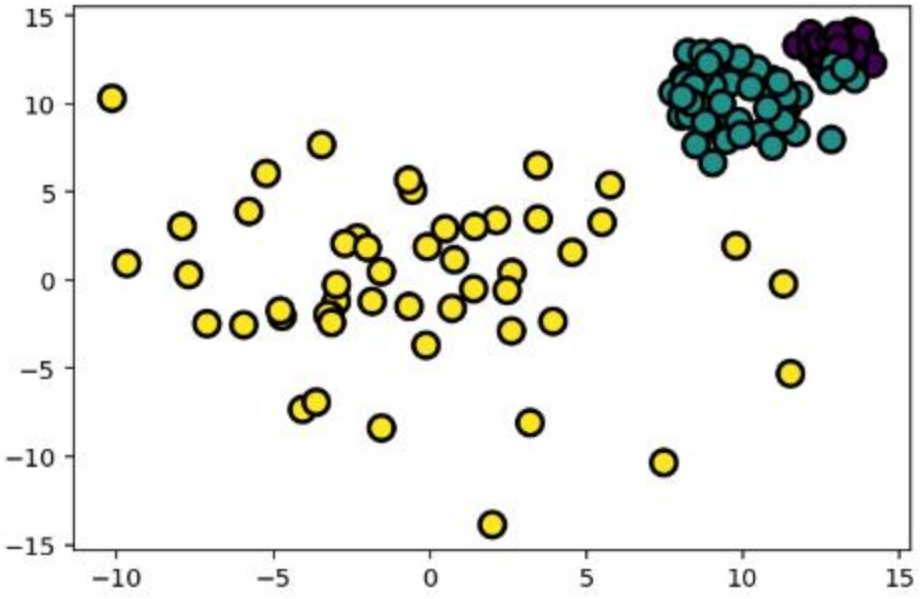


Original Points

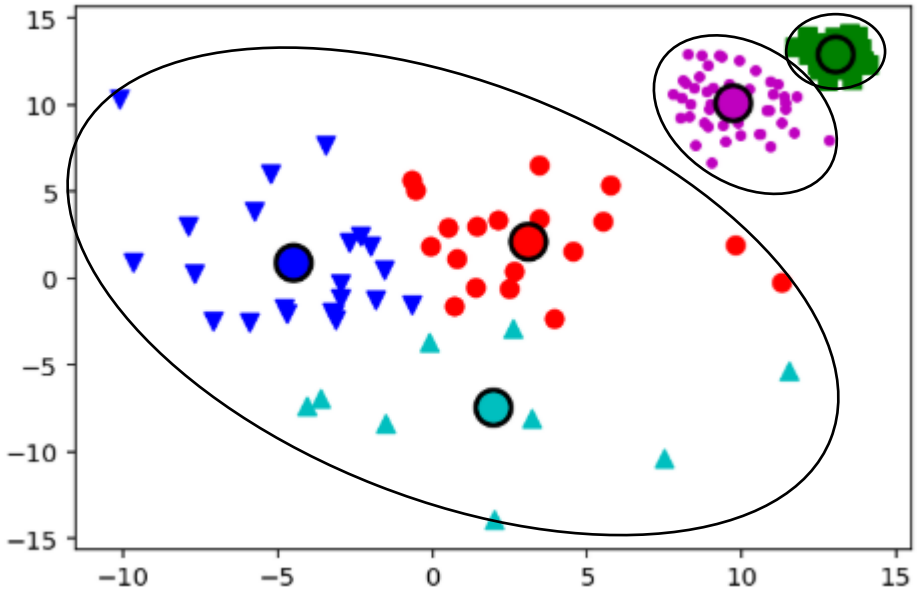


k-means (6 Clusters)

군의 밀도가 다를 경우 해결

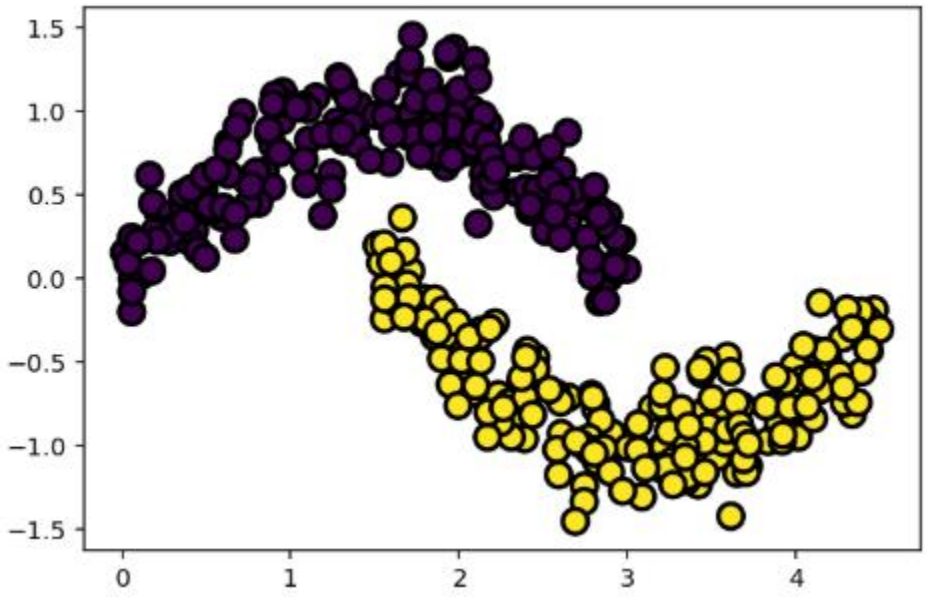


Original Points

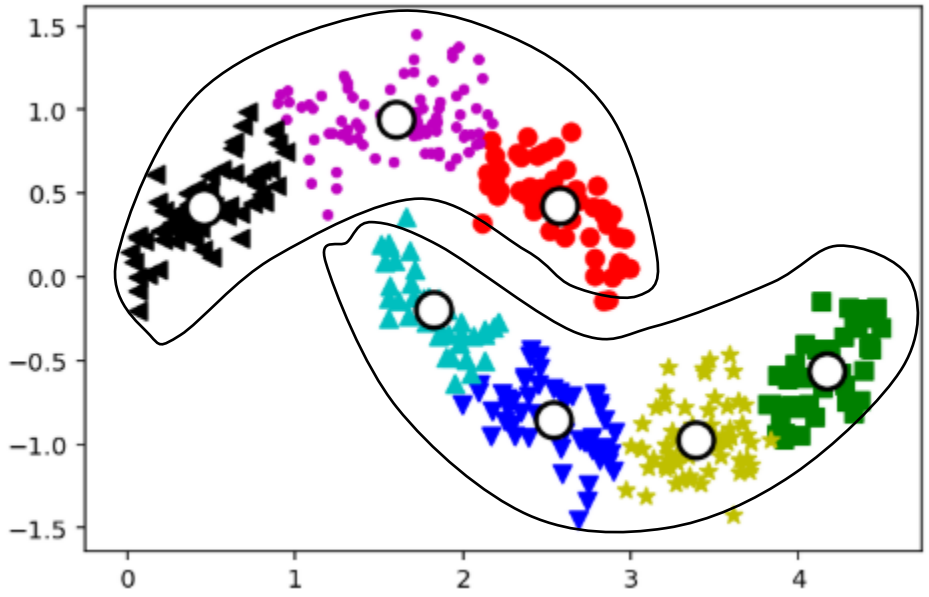


k-means (5 Clusters)

군이 구형이 아닌 경우 해결



Original Points



k-means (7 Clusters)