

# KNN(배포용)

2019년 1월 23일 수요일 오전 10:27

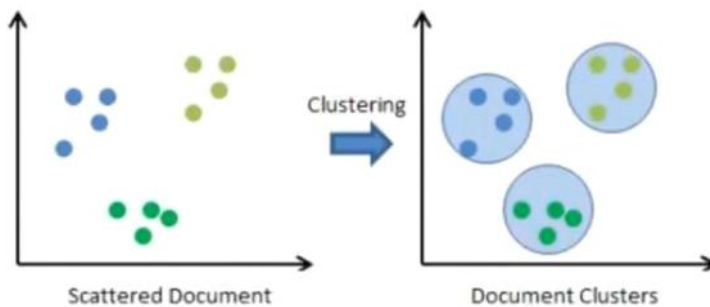
## 1. 분류의 2가지 종류

- Classification(지도 학습) : 의사결정나무, SVM 등.
- Clustering(비지도 학습)

## 2. Clustering

유사한 성격을 가진 개체를 묶어 그룹으로 구성하는 것.  
사전에 분류 라벨 필드가 없는 비지도 학습

아래 그림에서 확연히 구분되는 세 데이터 그룹 각각이 클러스터이다.

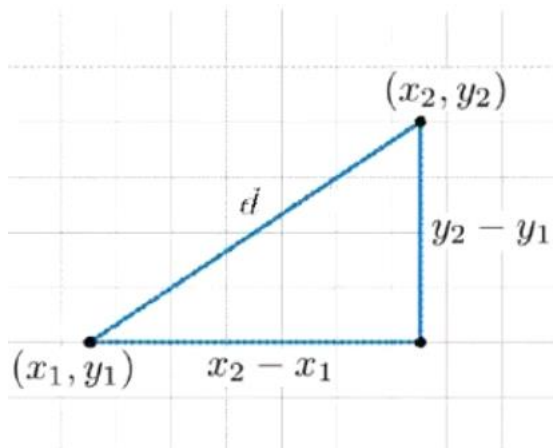


## 3. 군집분석의 원리

- 거리가 가까운(유사도가 높은) 개체들을 서로 묶어 하나의 그룹으로 정리
- 거리(유사도)의 계산 방법

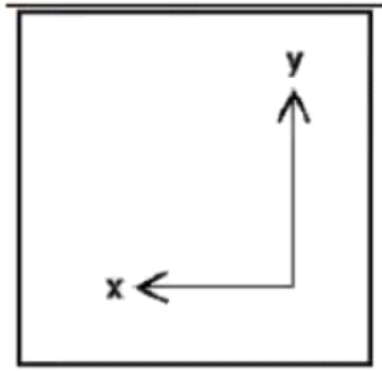
유클리드 거리(Euclidean distance)가 주로 사용됨

# 피타고라스 정리를 생각하면 됨.  $d$ 의 값을 구하기 위해서는 밑변제곱 더하기 높이제곱에 루트를 씌워서 계산하면 최단거리 계산이 된다.



맨해튼 거리(Manhattan distance)

# 그런데 거리를 무조건 가로질러 건너갈 수 없기 때문에 돌아갈 수 있는 걸리를 계산한 방식이 맨해튼 거리이다.



c. 군집화 방법

i. 계층적 군집화 방법  
Hierarchical method

ii. 비계층적 군집화 방법  
**K-means clustering, Unsupervised Learning**  
자기 조직화 지도(SOM, Self Organizing Maps) 등

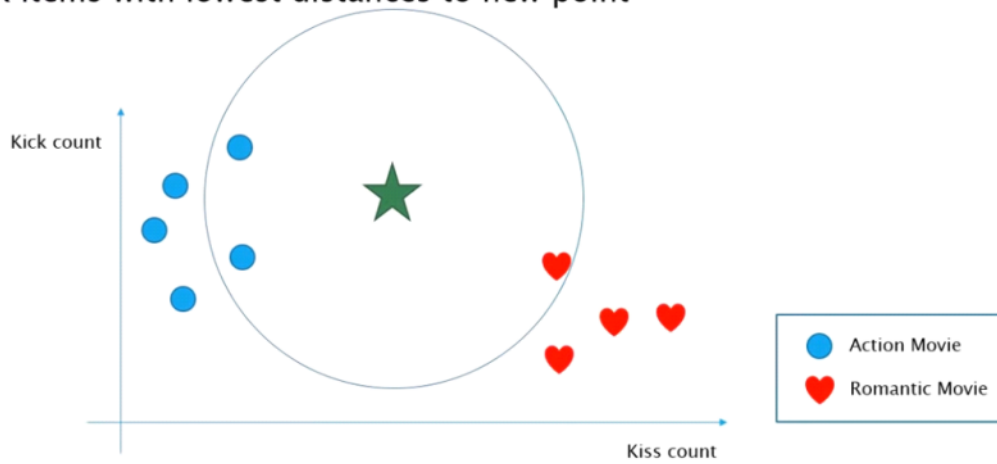
## How kNN (k - Nearest Neighbors) algorithm works



Assign K a value – preferably a small odd number  
(K = 3)



Find K items with lowest distances to new point  
(K = 3)



How do you find the distance between two points on a coordinate plane  
Using the Pythagorean Theorem?

