

<용어체크>

### 자율 학습

머신러닝의 학습 방법 중 하나로 정답 정보가 없는 상태에서 학습을 통해 모델을 만드는 것을 말하며 데이터가 어떻게 구성되었는지를 알아내는 문제 범주에 속한다.

### k-평균 알고리즘

주어진 데이터를 k개씩 클러스터로 묶는 알고리즘으로 각 클러스터와 거리 차이의 분산을 최소화 하는 방식이다. 자율 학습의 일종으로 EM 알고리즘을 이용한 클러스터링과 비슷한 구조를 가진다.

### GMM

가우시안 혼합 모델로 데이터의 분포가 가우시안 분포로 부터 생성되었다는 가정에서 시작하여 새로운 데이터가 주어질 때 데이터가 생성되었을 것으로 가정하는 가우시안 분포를 찾고자 하는 모델이다. 클러스터링 비지도학습으로 EM 알고리즘을 사용하여 학습하며 가능도 오차 함수를 목적함수로 사용한다.

<학습내용>

### 자율 학습

클러스터 분석과 K-평균 알고리즘

GMM 알고리즘

<학습목표>

자율 학습의 개념을 설명할 수 있다.

클러스터 분석을 이해하고 K-평균 알고리즘을 설명할 수 있다.

가우시안 혼합 모델을 이용한 군집화 알고리즘을 설명할 수 있다.

Q. K means 알고리즘과 KNN 알고리즘의 차이는 어떻게 될까요?

: KNN 알고리즘과 K Means 알고리즘은 비슷한 용어로 착각하기 쉬운 학습 알고리즘입니다. 일단 큰 차이점은 KNN 알고리즘은 지도 학습 알고리즘이고, K Means 알고리즘은 비지도 학습 알고리즘의 하나입니다. KNN은 데이터들을 거리 또는 유사도를 중심으로 묶어 분류를 해 주는 알고리즘이고 K Means 알고리즘은 주어진 클래스들이 없으므로 데이터를 분석하여 몇 개의 군집이 존재할지, 어떤 데이터들이 같은 성격을 갖는지를 파악할 수 있는 알고리즘입니다. 그러나 K Means 알고리즘의 경우 초기에 몇 개의 클러스터가 있는지를 결정해야 하며, 중심 값에 따라 클러스터 모양이 달라질 수 있는 단점을 가지고 있습니다. KNN은 거리 계산 방법에 따라 클래스의 분류가 달라질 수 있으므로 여러 번 수행하여 평균적인 값을 산출해야 할 수도 있습니다.

## 자율 학습

- ▶ 정답 정보가 없는 상태에서 학습을 통해 모델을 만드는 것이다.
- ▶ 데이터가 어떻게 구성되었는지를 알아내는 문제 범주에 속한다.
- ▶ 통계의 밀도 추정과 연관된다.
- ▶ 데이터의 주요 특징을 요약하고 설명한다.

## 클러스터 분석과 K-평균 알고리즘

- ▶ 그룹으로 만들 때 점들 사이가 어느 정도 떨어져 있는지에 대한 측정 지표이다.
- ▶ 자율 학습의 대표적인 접근 방법이다.
- ▶ 각 구간을 나눈 다음 중심을 찾고 중심을 기준으로 구간을 다시 나누고 변경 사항이 있을 경우 다시 중심을 찾아가는 방식이다.

## 가우시안 혼합 모델

- ▶ 가우시안 분포가 여러 개 혼합된 클러스터링 알고리즘이다.
- ▶ EM 알고리즘을 이용하여 클러스터링을 수행한다.
- ▶ Expectation step은 클러스터 별 부담율, 평균, 분산을 산출한다.
- ▶ Maximization step은 Expectation step에서 산출된 부담율을 최대화 하는 데이터 재분배를 수행한다.