

Part.05

Clustering

| Clustering이란

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

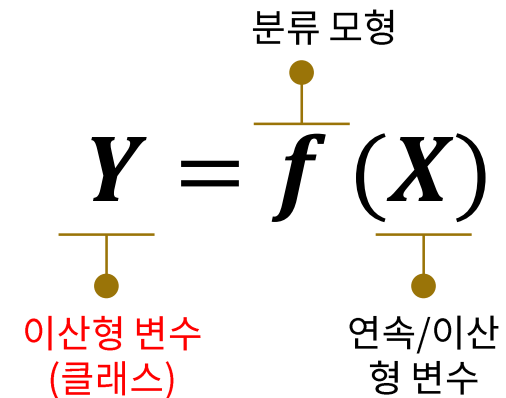
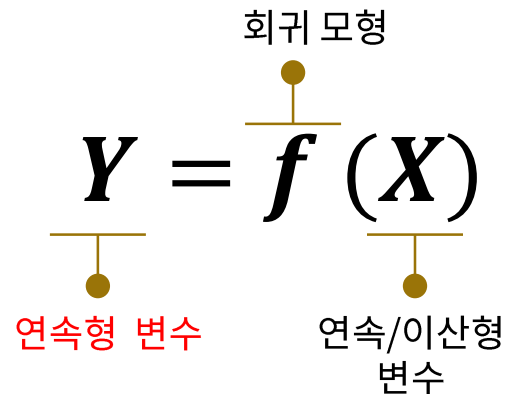
강사. 이경택

I Clustering이란

■ 지도 학습(supervised learning)

$Y = f(X)$ 에 대하여 입력 변수 (X)와 출력 변수 (Y) 의 관계에 대하여 모델링하는것
(Y 에 대하여 예측 또는 분류하는 문제)

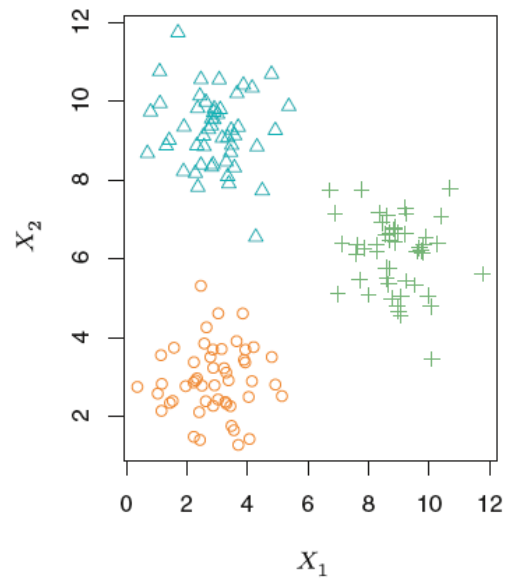
- 회귀 (regression): 입력 변수 X 에 대해서 연속형 출력 변수 Y 를 예측
- 분류 (classification): 입력 변수 X 에 대해서 이산형 출력 변수 Y (class)를 예측



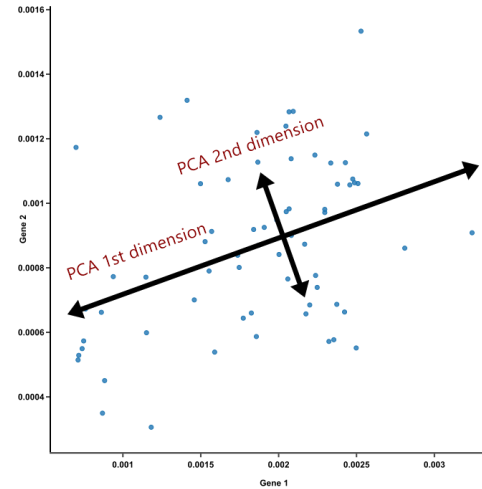
I Clustering이란

■ 비지도 학습(unsupervised learning)

- 출력 변수(Y)가 존재하지 않고, 입력 변수(X)간의 관계에 대해 모델링 하는 것
- 군집 분석 – 유사한 데이터끼리 그룹화
- PCA – 독립변수들의 차원을 축소화



군집 분석 예시



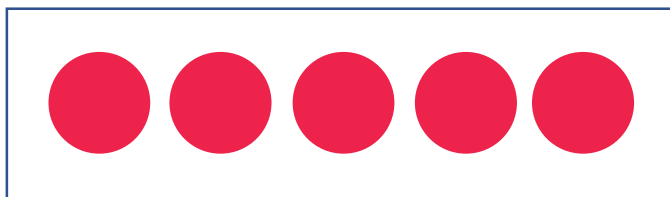
PCA 예시

I Clustering이란

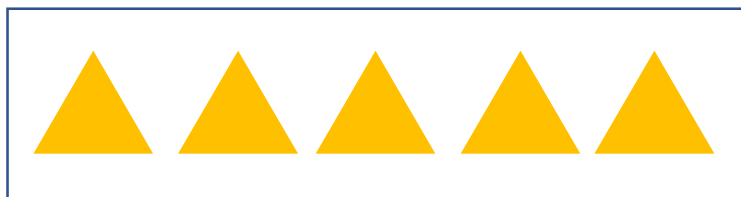
■ Clustering(군집분석)이란

- 각 데이터의 유사성을 측정하여 높은 대상 집단을 분류하고, 군집 간에 상이성을 규명하는 방법

전체데이터 (전체 뉴스기사)



군집1 (정치 관련 이슈)



군집2(스포츠 관련 이슈)



군집3(연예 관련 이슈)

- 고객 segmentation을 통한 마케팅 활용 방안 / 군집 별 추가 분석수행

I Clustering이란

- Clustering의 종류
 - K-means clustering : 데이터를 사용자가 지정한 k개의 군집으로 나눔
 - Hierarchical clustering (계층적 군집분석) : 나무 모양의 계층 구조를 형성해나가는 방법
 - DBSCAN : k개를 설정할 필요없이 군집화 할 수 있는 방법

Part.05

Clustering

| K-means clustering

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택