

파이썬 기반
빅데이터 처리
및 분석 기술





지난 시간에는

7

빅데이터 분류 분석

빅데이터 분류 분석의 절차

초모수와 모델 검증 방법

분류 심화: 나이브 베이즈 기법





이번 시간에는

8

빅데이터 군집 분석

빅데이터 비지도 학습의 개념

심화: K-평균 군집 분석

심화: 가우스 혼합 모델





이번 시간에는

8

빅데이터 군집 분석

학습 목표

- 비지도 학습과 군집 분석, 차원 축소의 개념을 이해한다.
- K-평균 군집 분석을 실습한다.
- 가우스 혼합 모델을 실습한다.





빅데이터 군집 분석

1

빅데이터 비지도 학습의 개념

머신러닝

지도학습

레이블 데이터가 있는 것

비지도 학습

레이블 데이터가 없는 것





8-1 빅데이터 비지도 학습의 개념



대구가톨릭대학교
사물인터넷(IoT)과 함께하는 빅데이터

빅데이터 군집 분석



1

빅데이터 비지도 학습의 개념

비지도 학습의 개념

군집 분석과 차원 축소





1. 비지도 학습

1) 개요

비지도 학습

- 머신러닝의 일종
- 데이터 어떻게 구성되었는지를 알아내는 문제의 범주에 속함
- 지도 학습과는 다르게 레이블이 없거나 참조하지 않음

데이터 세트가 스스로 데이터를 설명하게 하는 것





1. 비지도 학습

1) 개요

◆ 데이터 특징과 레이블

비지도 학습

군집화

구체적인 특성을
공유하는 군집을 찾는 것

차원 축소

데이터 세트의
특징행렬을 간결하게
표현하는 방법

특징행렬의
차원이 클수록
차원 축소의 효용이 높음

특징

레이블

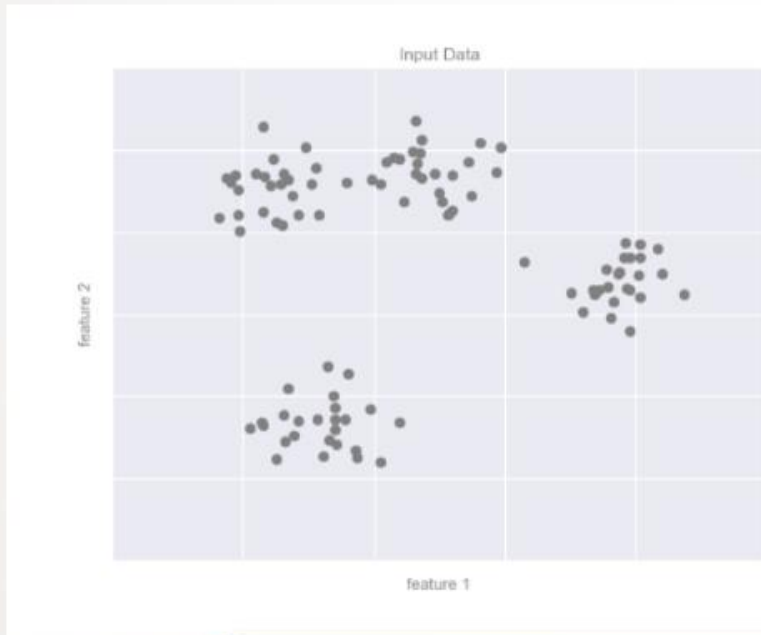
기온		습도	
20		60	
30		80	
30		75	



1. 비지도 학습

2) 군집화

◆ 예제



특징 데이터를 이용하여 레이블을 예측하는 모델 사용 불가



1. 비지도 학습

2) 군집화

◆ 예제



레이블이 없는 데이터를 유사한 성질을 갖는 것들끼리 묶는 알고리즘



1. 비지도 학습

2) 군집화

◆ 장점

- 각각의 그룹의 성질을 분석하고 의미를 부여하여 레이블을 새로 할당 할 수 있음
- 사전에 할당된 레이블이 있는 경우라도 적용 가능
- 군집분석을 통해 동일하게 묶인 데이터 세트로 지도 학습 수행도 가능



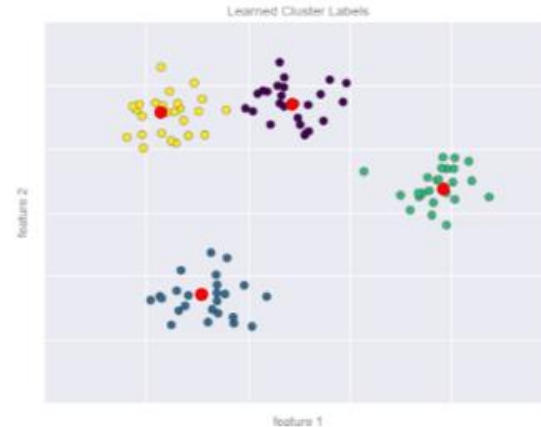


1. 비지도 학습

2) 군집화

◆ K-평균 알고리즘(k-means)

- K개의 군집 중심으로 구성된 모델에 데이터를 적합시키는 방법
- 할당된 중심으로부터 각 점까지의 거리를 최소화하는 점이 최적의 중심점
- 대량의 복잡한 데이터 세트에서 적용하면 유용하게 정보 추출 가능





1. 비지도 학습



3) 차원 축소

◆ 개요

- 레이블이나 다른 정보를 데이터 세트 자체의 구조로부터 추론하는 방식
- 전체 데이터 세트의 품질을 떨어뜨리지 않으면서 차원의 개수를 줄이는 표현 방식을 찾아내는 방법





8-1 빅데이터 비지도 학습의 개념



대구가톨릭대학교

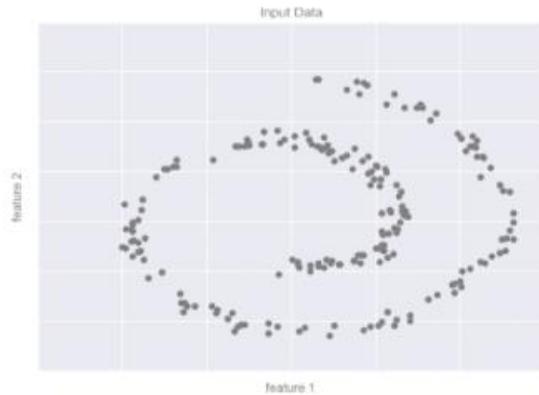
사물인터넷(IoT)과 함께하는 빅데이터

1. 비지도 학습



3) 차원 축소

◆ 레이블 없는 데이터의 구조 추론

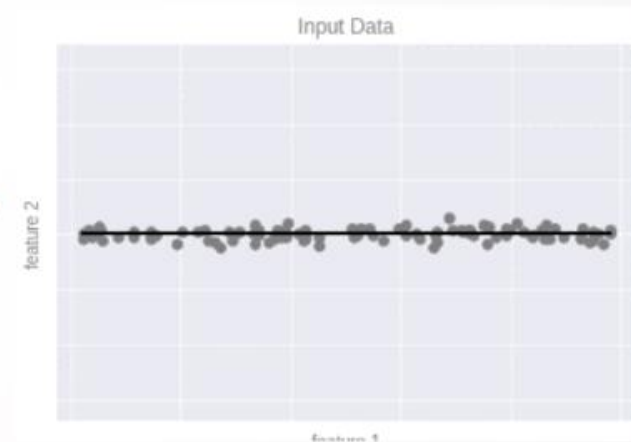




1. 비지도 학습

3) 차원 축소

◆ 레이블 없는 데이터의 구조 추론



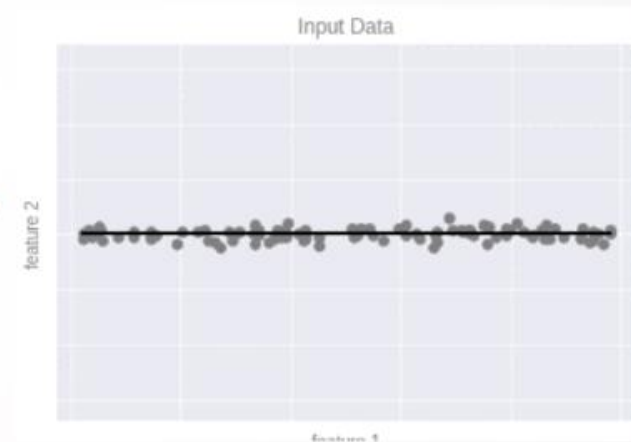
2차원 → 1차원



1. 비지도 학습

3) 차원 축소

◆ 레이블 없는 데이터의 구조 추론



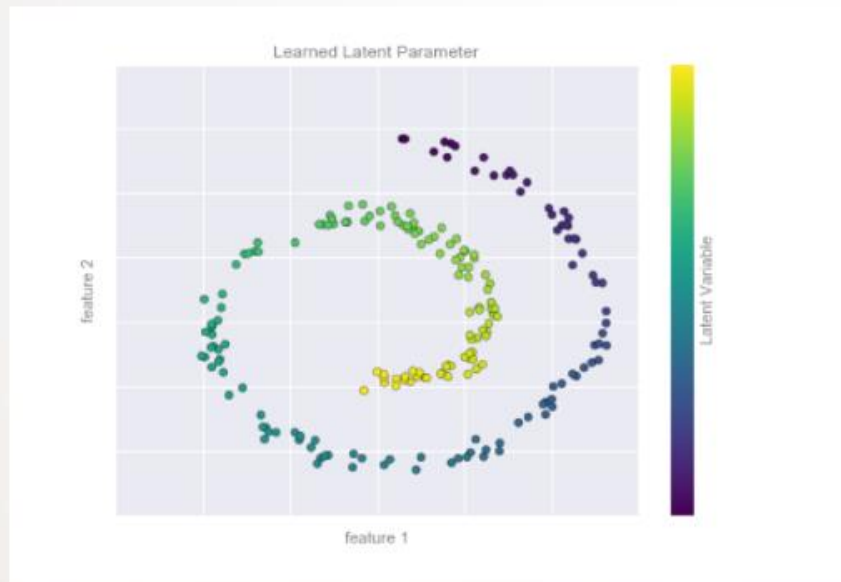
차원의 수를 줄여도 데이터의 구조는 변하지 않음



1. 비지도 학습

3) 차원 축소

◆ 등거리 사상 알고리즘을 이용한 시각화



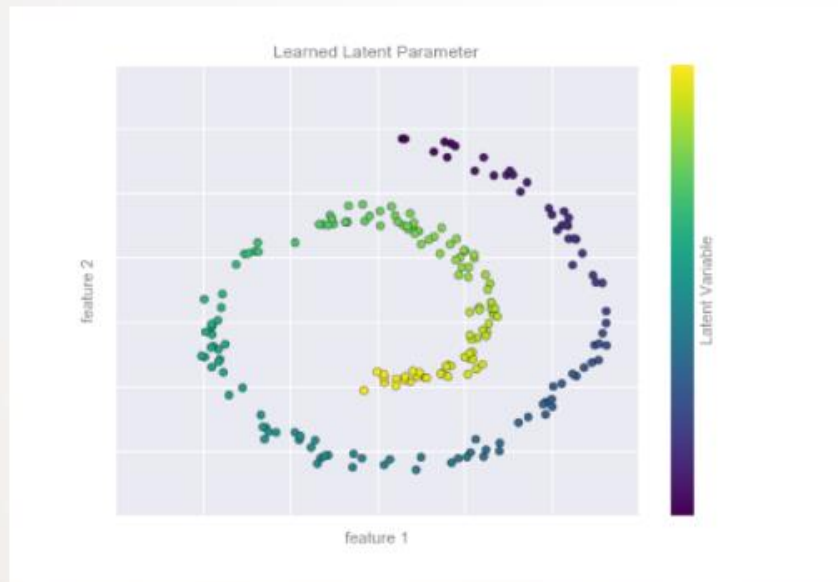
차원 축소는 고차원 데이터일수록 효과적



1. 비지도 학습

3) 차원 축소

◆ 등거리 사상 알고리즘을 이용한 시각화



차원 축소는 머신 러닝의 결과를 높이기 위한 보조 수단으로 사용



이번 시간에는

1

빅데이터 비지도 학습의 개념

비지도 학습의 개념

군집 분석과 차원 축소





이번 시간에는

실습 참고 자료

- Colab 노트북 파일
- Scikit-Learn 공식 사이트 자료
- https://scikit-learn.org/stable/user_guide.html





다음 시간에는



2

심화: K-평균 군집 분석

K-평균 소개

실습: 숫자 데이터 분석

