

파이썬 기반의  
빅데이터 처리 및  
분석 기술





# 지난 시간에는

5

빅데이터 시각화

Matplotlib 소개와 사용법

선그림과 점그림

다중 서브 플롯



## 이번 시간에는

6

빅데이터 회귀 분석

### 머신 러닝

- 인공지능의 한 분야
- 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야
- 표현(데이터 평가)과  
일반화(아직 알 수 없는 데이터에 대한 처리)



# 이번 시간에는

6

빅데이터 회귀 분석

머신러닝의 개념과 기본 절차

Scikit-Learn API 사용법

회귀 분석 심화



# 이번 시간에는

6

빅데이터 회귀 분석

## 학습 목표

- 머신러닝의 개념과 절차에 대해 이해한다.
- Scikit-Learn API의 사용법을 이해한다.
- Scikit-Learn API를 사용하여 회귀 분석을 진행한다.



# 빅데이터 회귀 분석

1

머신러닝의 개념과 기본 절차

머신러닝의 정의

머신러닝의 범주



## 1. 머신러닝

### 1) 개요

#### “머신러닝의 정의”

- 인공지능의 하위 분야로 분류
- 데이터 모델을 구축하는 체계적인 수단
- 데이터를 통한 학습 모델을 도출하고 새로운 데이터에 대한 추론과 패턴 예측에 활용





1. 머신러닝



1) 개요

“머신러닝의 범주”

지도학습  
(Supervised learning)

데이터의 측정된 특징과  
데이터와 관련된 레이블  
사이의 관계를 모델링

비지도 학습  
(Unsupervised learning)

레이블을 참조하지 않고  
데이터 세트의 특징을 모델링





1. 머신러닝

1) 개요

“지도학습의 예”

특징		레이블
기온	습도	불쾌지수
20	60	10
30	80	80
30	70	70
35	75	?

- 데이터 특징과 레이블 사이의 관계 모델링



1. 머신러닝



1) 개요

“비지도학습의 예”

특징	
기온	습도
20	60
30	80
30	70

- 데이터 세트의 특징을 모델링



1. 머신러닝

2) 회귀 분석

회귀 분석

레이블이 연속적인  
수치형 데이터일 경우

분류 분석

레이블이 범주, 클래스 등  
이산적 데이터일 경우



1. 머신러닝



2) 회귀 분석

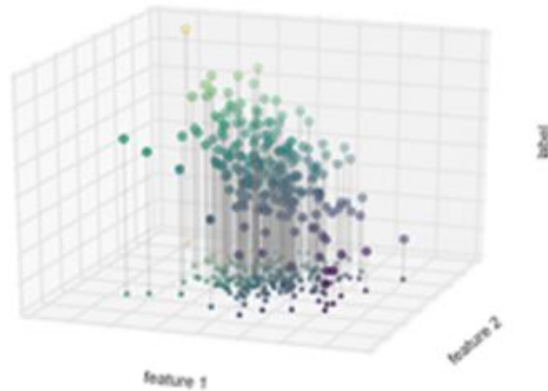


2차원 데이터 : 데이터를 설명하는 특징 2개



1. 머신러닝

2) 회귀 분석



3차원 데이터: 색상을 높이로 치환



## 6-1 빅데이터 회귀 분석



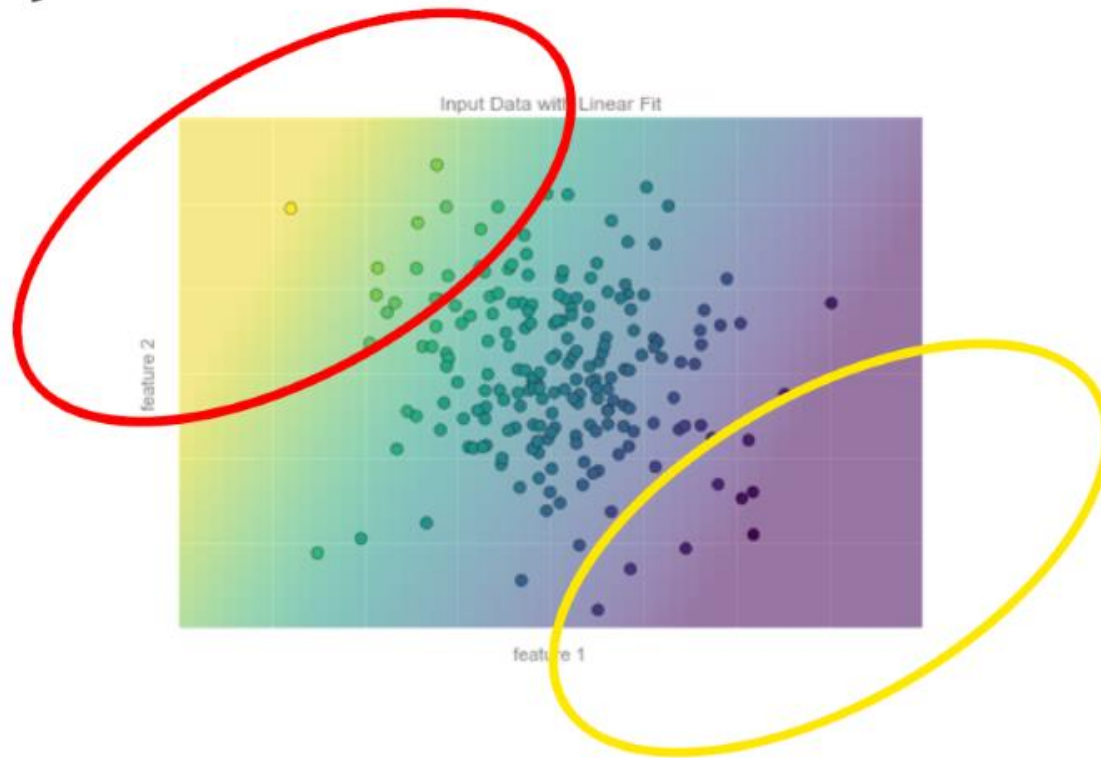
대구가톨릭대학교

사물인터넷(IoT)과 함께하는 빅데이터

### 1. 머신러닝



### 2) 회귀 분석





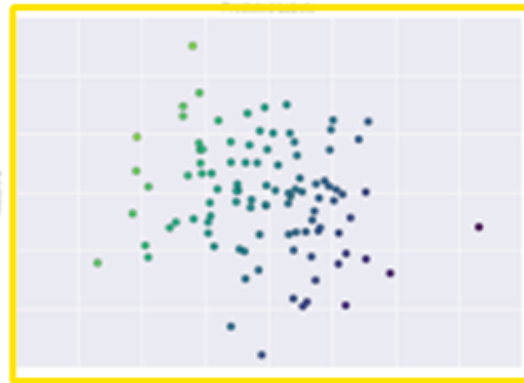
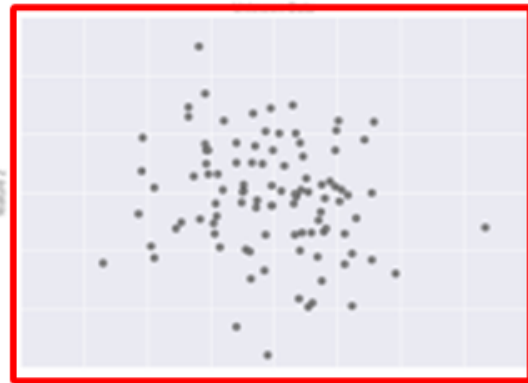


1. 머신러닝



2) 회귀 분석

“회귀 분석 결과 활용 방법”



천문학에서 활용 → 측광 적색편이

새로운 데이터 레이블을 조사하기 위한 비용과 시간 절감 효과



# 이번 시간에는

1

머신러닝의 기본 개념

머신러닝의 정의

머신러닝의 범주



## 이번 시간에는

### 실습 참고 자료

- Colab 노트북 파일
- Matplotlib 공식 사이트  
→ <https://matplotlib.org/tutorials/index.html>





다음 시간에는

2

Scikit Learn API 사용법

Scikit Learn API 장점

Scikit Learn 기본 사용법

간단한 회귀 분석