



예측모델 구현 기술



회귀분석 및 시각화



예측이란 무엇인가?

◆ 예측(Prediction, Forecasting)

미리 헤아려 짐작함

- '짐작'의 정확도를 높이기 위해선 '헤아림'에 적합한 데이터가 필요함

예측은 과거의 또는 확보한 정보를 분석해 패턴을 찾고
그 패턴을 바탕으로 앞으로의 흐름 또는 연관성을 찾아내는 것임

회귀분석 및 시각화



학습이란 무엇인가?

◆ 학습(Learning)

01

기계가 학습한다

- = 데이터에서 패턴을 찾아낸다

02

학습을 완료했다

- = 주어진 데이터를 대표하는 가장 좋은 패턴을 찾았다

회귀분석 및 시각화




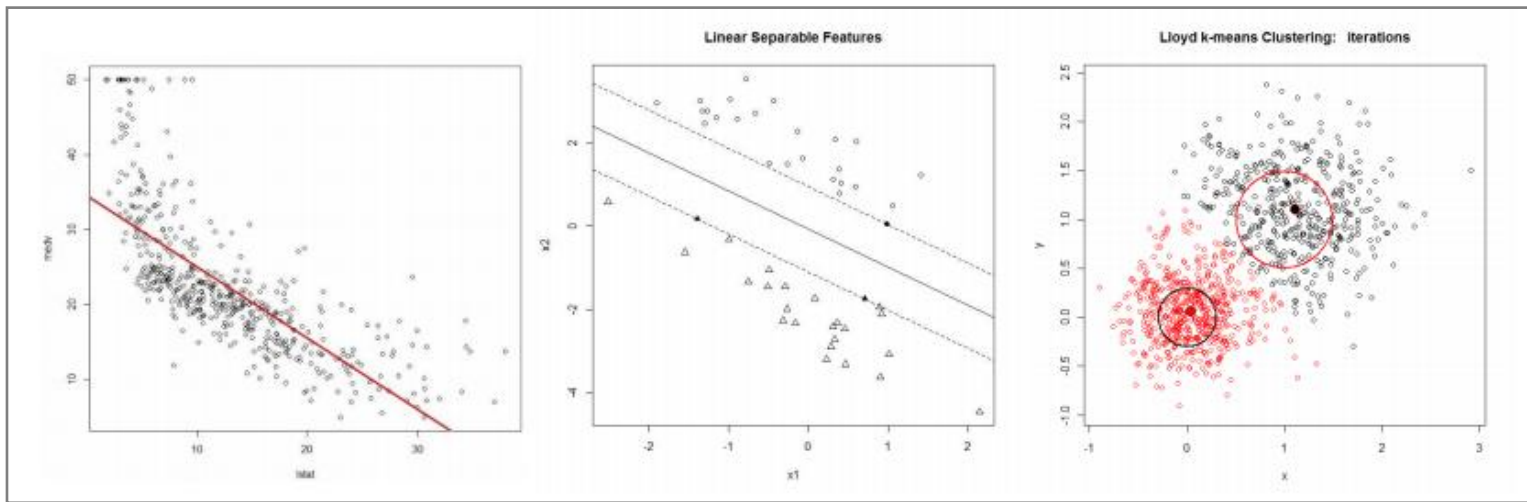
학습이란 무엇인가?

◆ 학습(Learning)

03

학습을 필요한 데이터를 모으는 방법

- 모아놓고 시작 : 인터넷에서 다운로드, 자동화 수집 도구 사용
- 하면서 모으기  시뮬레이션을 하면서 모으기
예 게임



회귀분석 및 시각화



학습 유형

◆ 기계 학습(Machine Learning)

01 지도학습

- 정답지(LABEL)가 있는 데이터에서 일정한 패턴을 찾는 학습

02 지도학습

- 정답지(LABEL)가 없는 데이터에서 일정한 패턴을 찾는 학습

예 군집분석

03 강화학습

- 보상(Reward)을 기반으로 하는 학습

회귀분석 및 시각화



학습 목적

회귀분석

- 앞으로 어떻게 변화할 것인가?

분류분석/군집분석(군집화, Clustering)

- 이 데이터는 어디에 속하지?

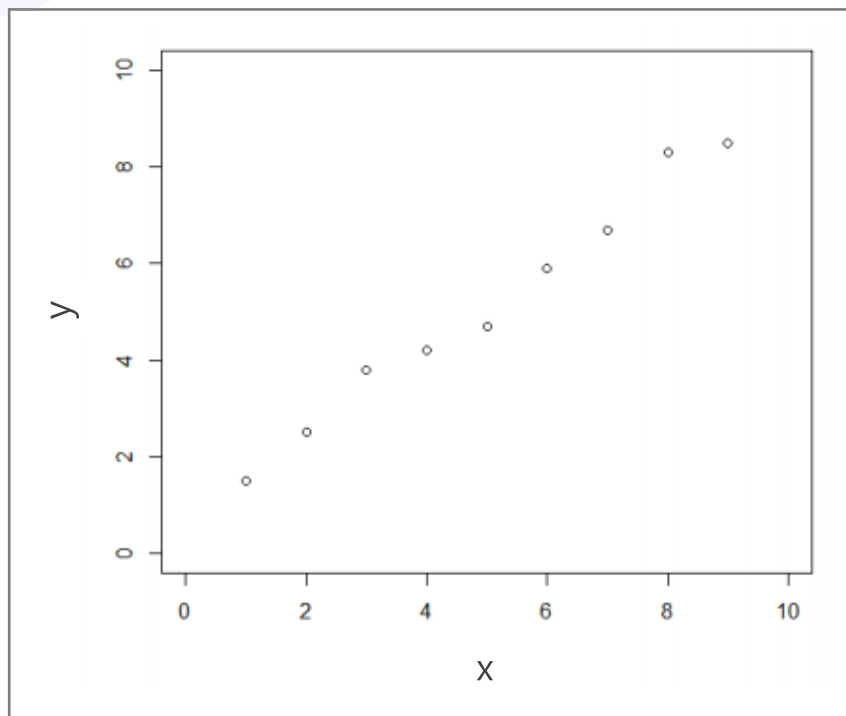
회귀분석 및 시각화



선형 회귀(Linear Regression)

종속 변수 y 와 한 개 이상의 독립 변수 x 와의 선형 상관 관계를 모델링

오차가 가장 적은 $y = f(x)$ 함수를 찾는 문제(최소자승법)



■ $f(x) = w x + b$

- w : 기울기(가중치)
- b : 절편



과목명	데이터시각화	주차명	14. 실데이터 활용, 문제해결 (2) 예측과 분류 모델 구현 기술	페이지번호	02_02_14
학습목차	<div>16150856_14_02.mp4</div> <div><div></div></div> <div><div></div><div></div><div></div></div> <div><div>▶</div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div><div></div></div> <div><div></div><div></div></div>				

시계열 분석 및 시각화



시계열(Time-series) 데이터

01

시간 순서대로 인덱스 된 연속된 데이터 점으로 이루어진 데이터

02

시간 경과에 따라 순서대로 관측되는 데이터



예 연도별, 분기별, 월별, 일별, 시간별 등

03

일반적인 머신러닝 모델링과 다른 접근 방식을 적용해야 함

- 경향(Trend)
- 주기(Cycle)
- 계절성(Seasonality)
- 불규칙성(Irregular)

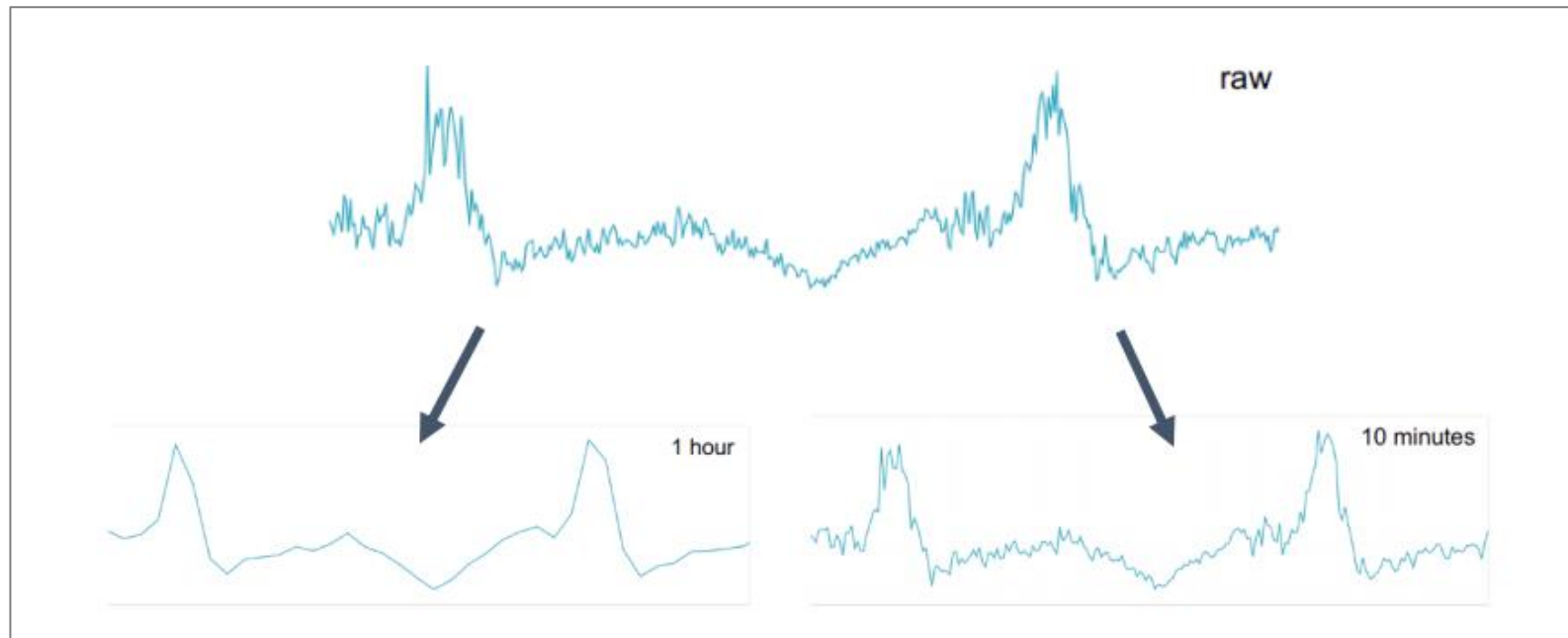
시계열 분석 및 시각화



시계열(Time-series) 데이터 분석

01 과거와 현재의 분석은 정확하게 수행이 가능함

02 대상의 특징과 다른 요소와의 관계 규명에 유리함



시계열 분석 및 시각화



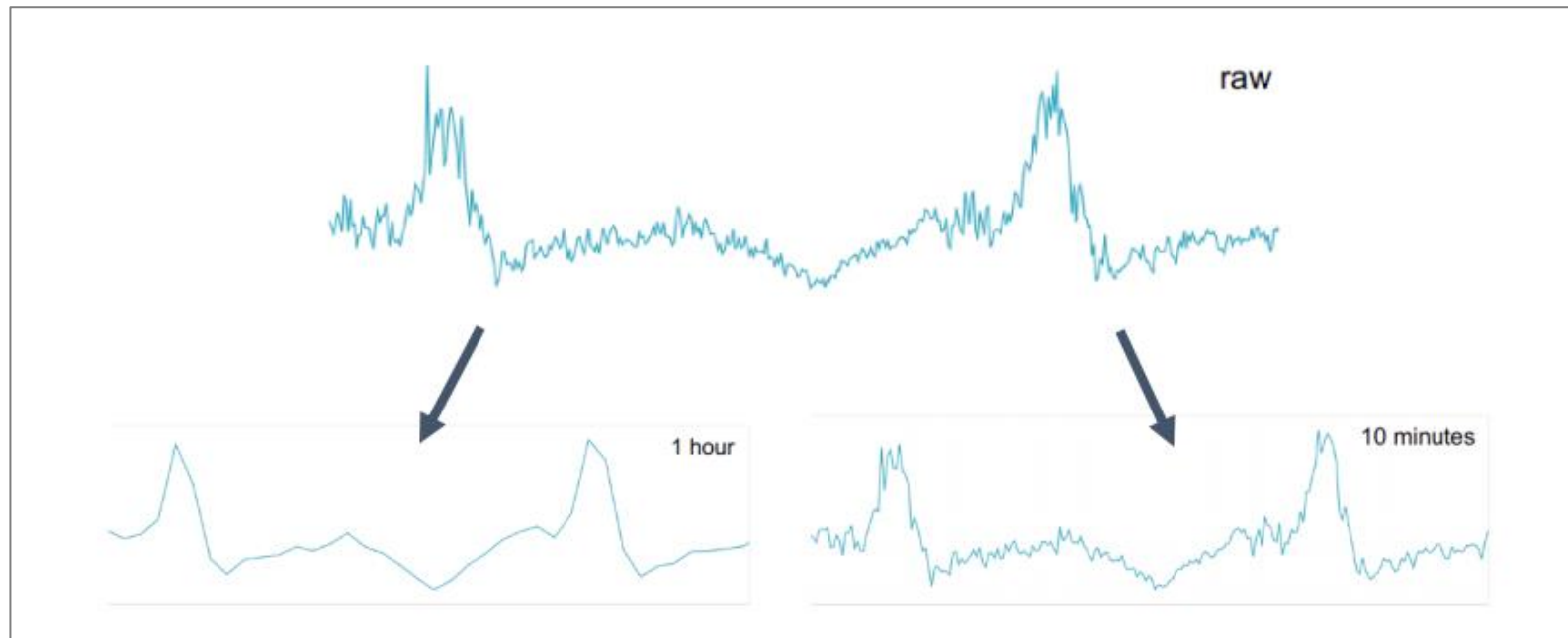
시계열(Time-series) 데이터 분석

03

시계열 분석은 미래 예측을 위한 도구가 아님

예

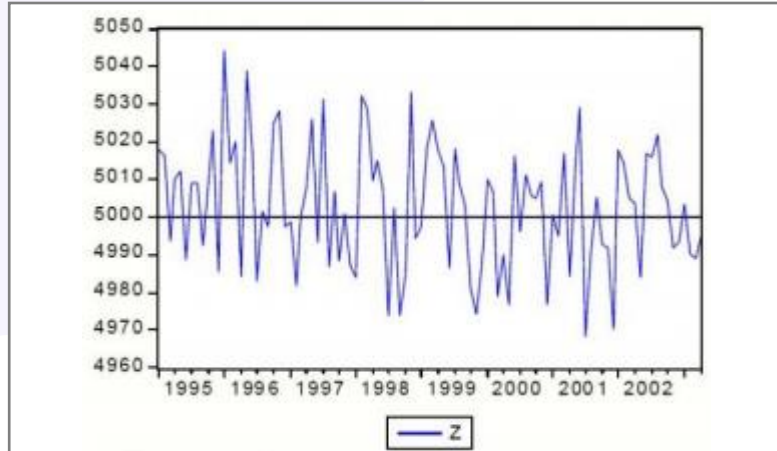
주식, 부동산 예측



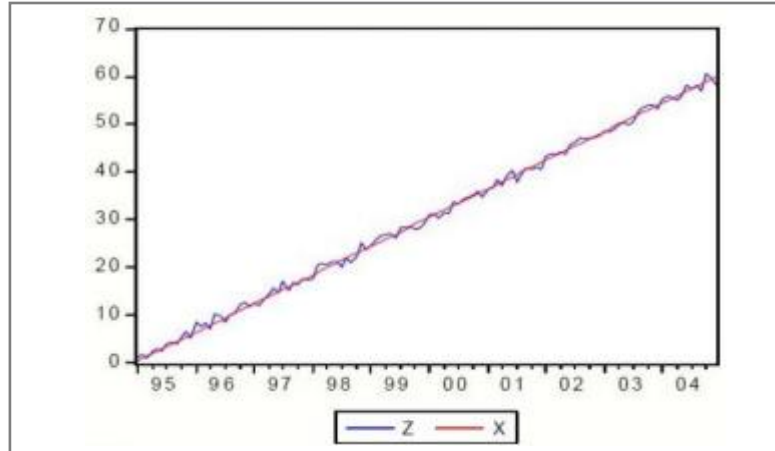
시계열 분석 및 시각화



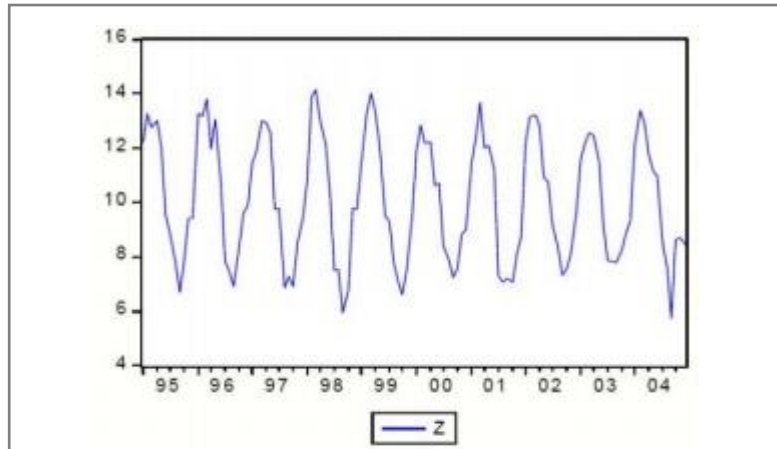
시계열(Time-series) 데이터 형태



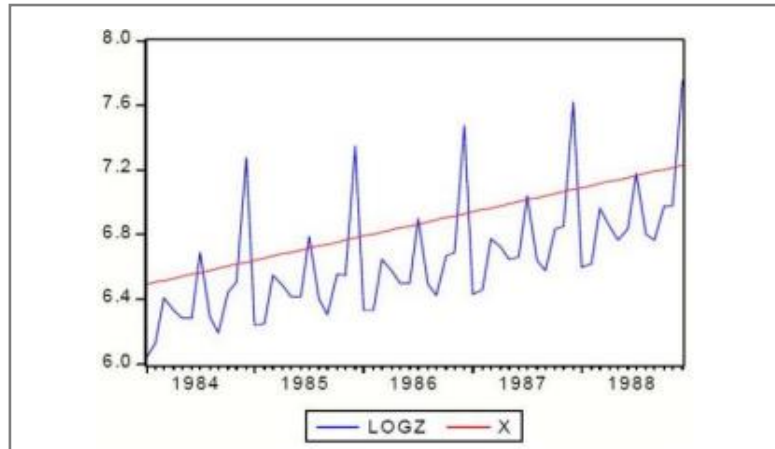
불규칙변동



규칙변동



계절변동



추세 + 계절변동

시계열 분석 및 시각화

시계열(Time-series) 분석



- 이전의 결과와 이후의 결과 사이에서 발생하는 자기상관성
- 이전에 생긴 불규칙한 사건이 이후의 결과에 편향성을 초래하는 이동평균

시계열 분석 및 시각화



시계열(Time-series) 분석

AR 모형

- 자기상관(Autocorrelation)을 시계열 모형으로 구성한 모형은 변화

MR 모형

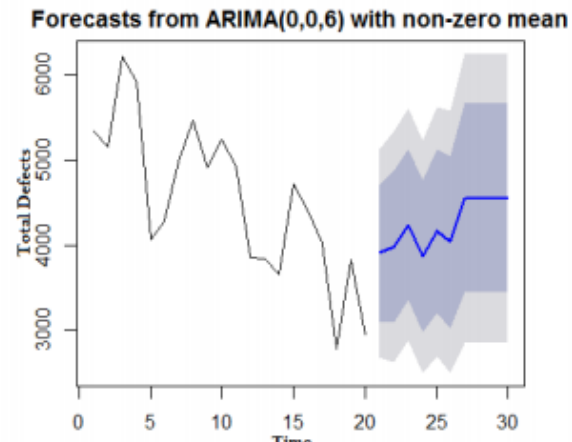
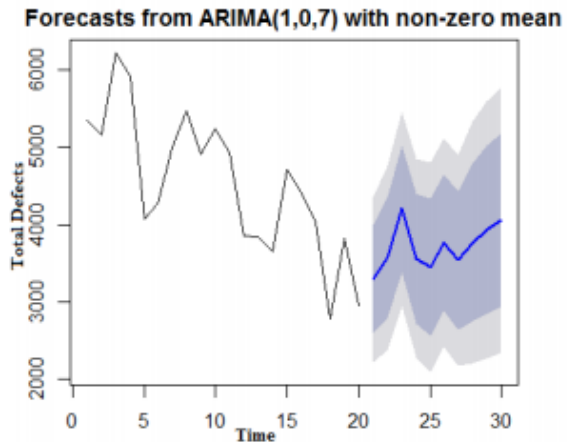
- 이동평균(Moving Average)을 시계열 모형으로 구성한 모형은 변화

시계열 분석 및 시각화

시계열(Time-series) 분석

ARIMA 모형

- Auto-regressive Integrated Moving Average
- 과거의 관측값과 오차를 사용해 현재의 시계열값을 설명하는 ARMA 모형을 일반화 한 모형
- 시계열 데이터는 반드시 '안정된 시계열' 형태여야 함





분류 모델 구현 기술



의사결정나무 분석 및 시각화



의사결정나무(Decision Tree)

지도학습 ML중 분류(Classification) 알고리즘 중 하나로
ML 알고리즘 중 직관적으로 이해하기 쉬운 알고리즘임

데이터내의 규칙을 학습을 통해 자동으로 찾아내
나무 모양의 분류 규칙을 만들

의사결정에 영향을 미치는 독립변수(Feature)에 대한
영향력을 분류하거나 예측함



의사결정나무 분석 및 시각화



의사결정나무(Decision Tree)

◆ 목표 변수(해석대상)

01

이산형인 경우 분류나무(Classification Tree)

02

연속형인 경우 회귀나무(Regression Tree)

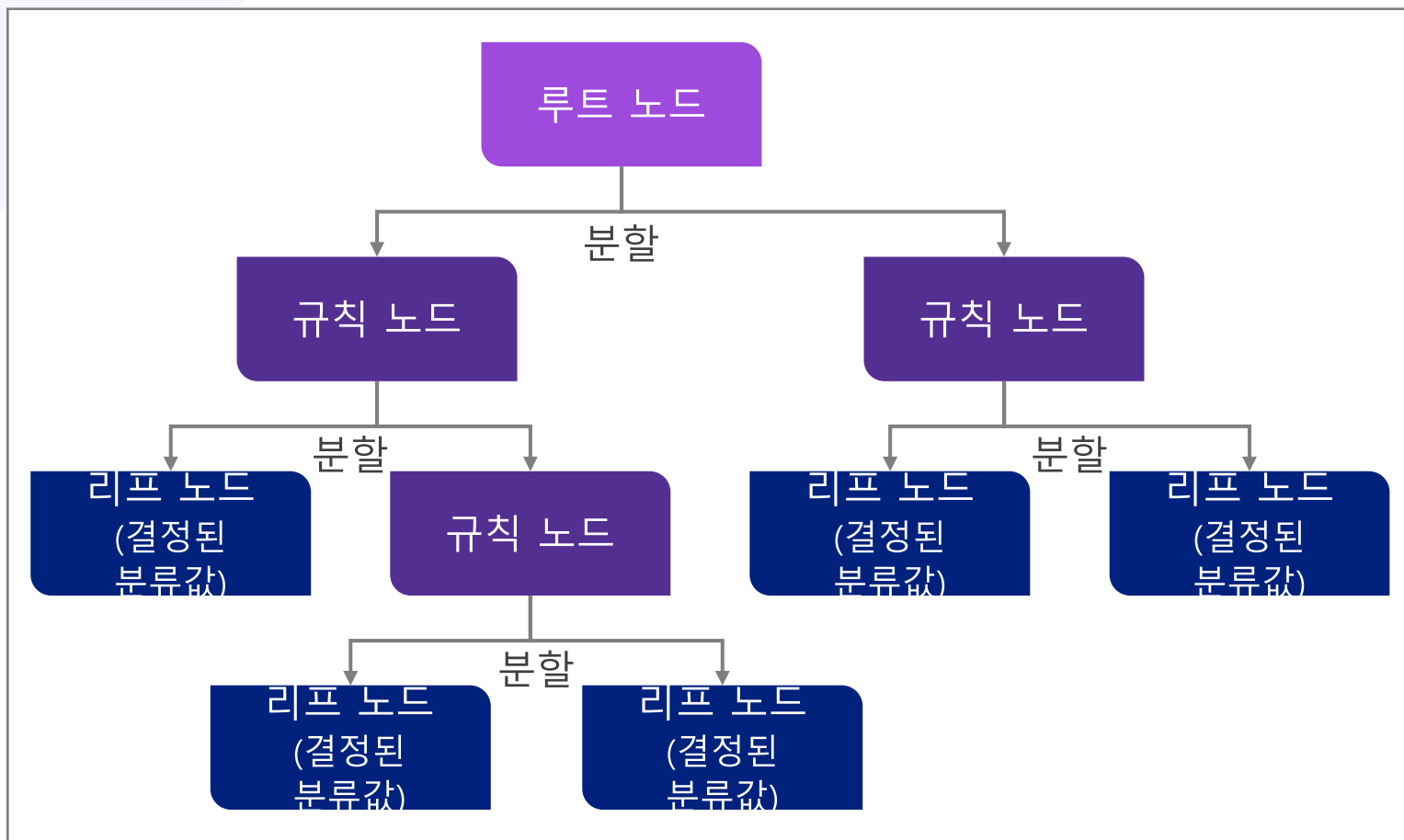


의사결정나무 분석 및 시각화



의사결정나무(Decision Tree)

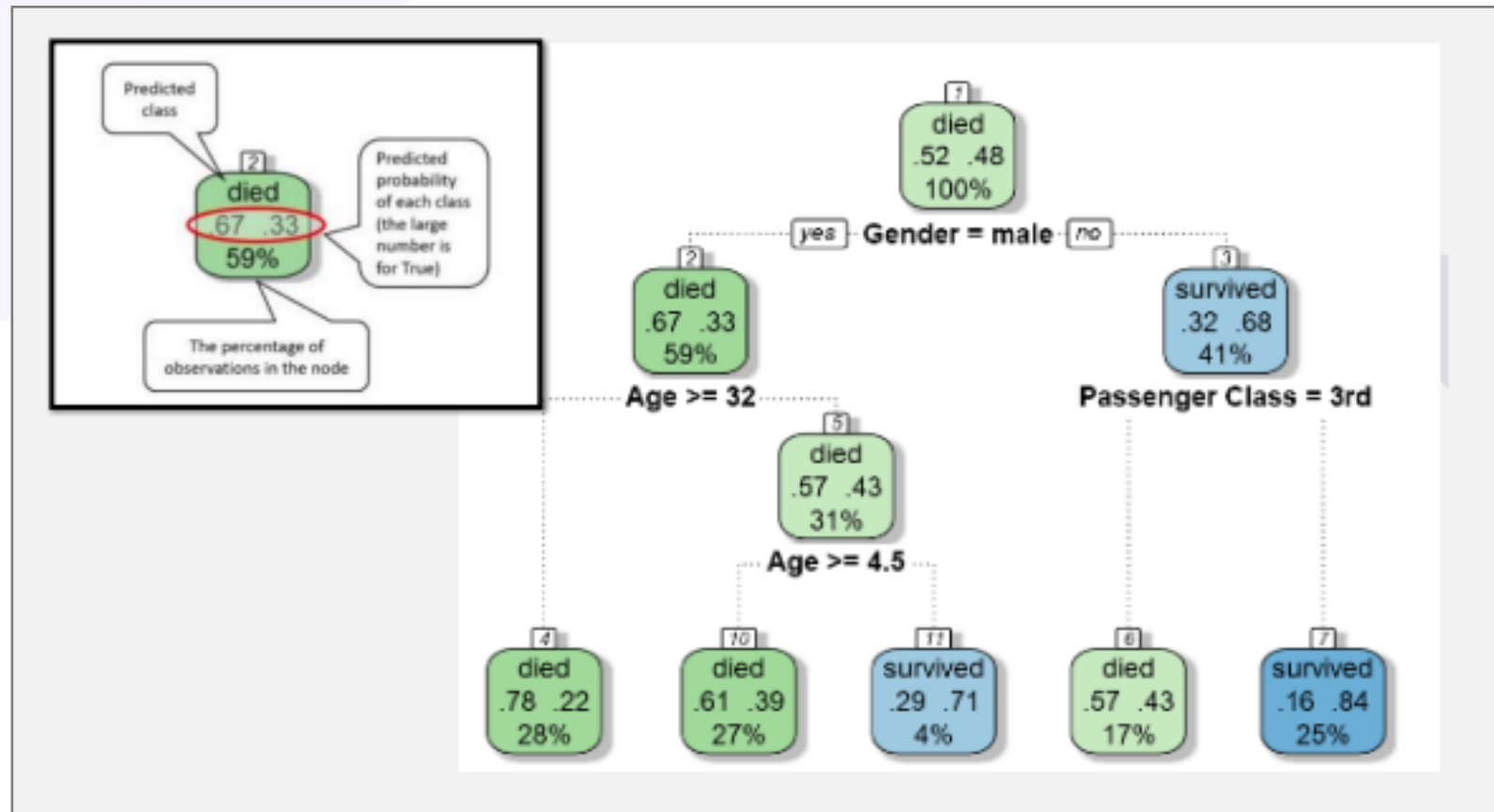
의사결정나무 구조와 용어



의사결정나무 분석 및 시각화



의사결정나무(Decision Tree)



의사결정나무 분석 및 시각화



의사결정나무(Decision Tree)

타이타닉 생존율 분석 예제 실습

Target Variable

Survived

Input Variables

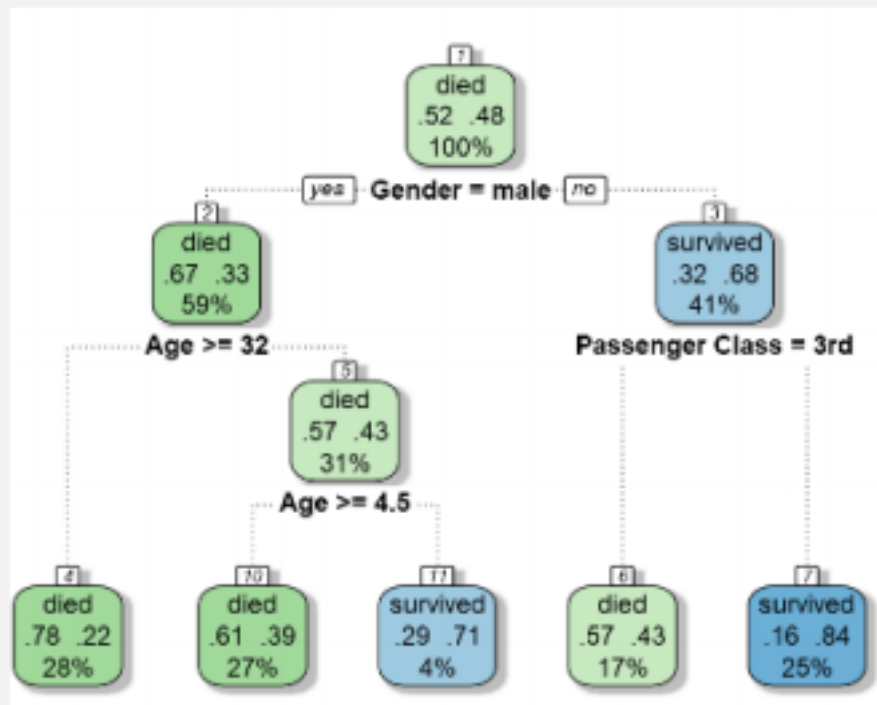
Age

Sex

Pclass

Fare

Cabin



↑ 순서는 상관 없음

의사결정나무 분석 및 시각화



의사결정나무(Decision Tree)

장점	단점
쉽고 직관적 알고리즘 ➡ 의사결정을 위한 규칙생성으로 해석 및 적용이 쉬움	데이터의 공간을 입력 변수의 수직 축으로만 나눔
독립변수(입력변수)의 스케일링이나 정규화 등의 사전 가공 영향도가 크지 않음	과접합(Overfitting)으로 인한 성능 저하 ➡ 이를 극복하기 위한 튜닝 필요함
명목형 변수, 연속형 변수 모두 처리 가능	한 번 발생한 분기는 다시 되돌릴 수 없음

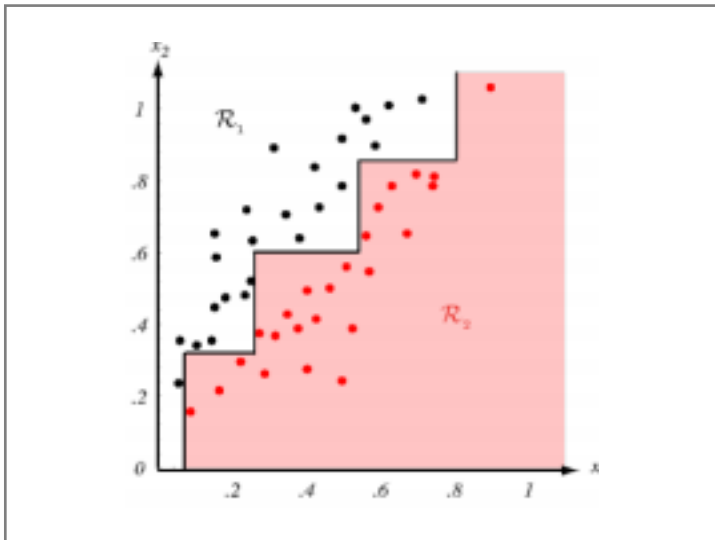
의사결정나무 분석 및 시각화



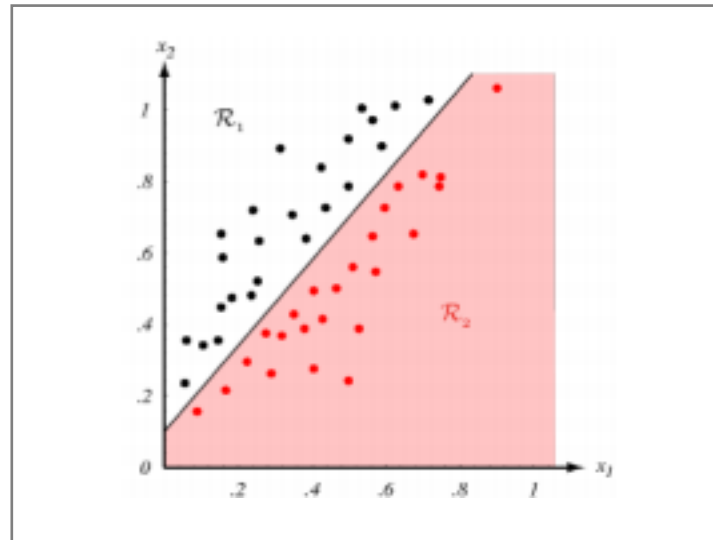
의사결정나무(Decision Tree)

단점 추가

데이터의 공간을 입력 변수의 수직 축으로만 나눔



Possible



Impossible

의사결정나무 분석 및 시각화



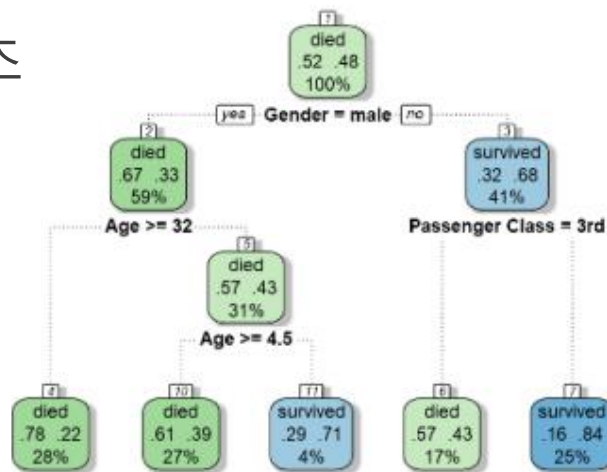
의사결정나무(Decision Tree)

실습 옵션

Tree Parameters

- Maximum depth
 - 노드의 깊이
 - 숫자가 커지면 시간도 많이 걸리고 과적합이 발생됨
- Minimum bucket Size
 - 마지막 노드가 되기 위한 최소 관측 값
 - 숫자가 커지면 단순한 트리 구조

Advanced Parameters



의사결정나무 분석 및 시각화



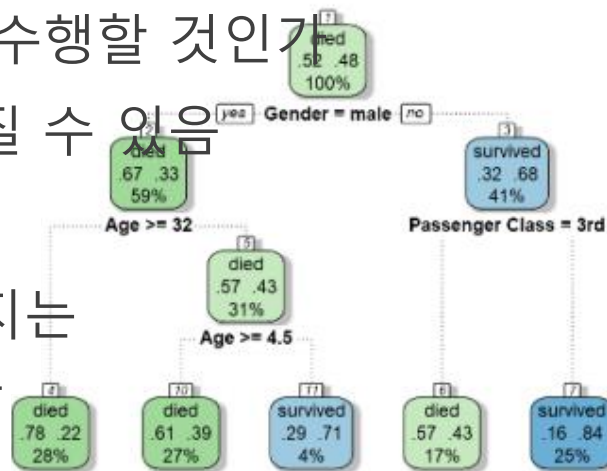
의사결정나무(Decision Tree)

실습 옵션

Tree Parameters

Advanced Parameters

- Complexity
 - 더 이상 분할되지 않은 복잡도의 정도
 - 이 수치가 낮아지면 정확도는 좋아지나 과적합 문제가 발생
- Cross-validation
 - 데이터셋을 몇 번 구간을 나눠 수행할 것인가
 - 값이 올라가면 오류 값이 낮아질 수 있음
- Maximum attempts
 - 하나 이상의 최적의 깊이를 가지는 나무구조를 구성하는 시도 횟수



군집분석 및 시각화

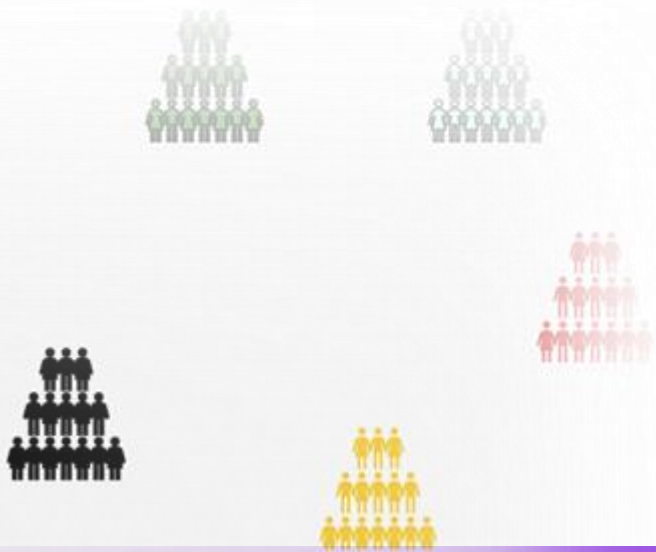


군집분석(Clustering)

클러스터링은 데이터에서 비슷한 객체(데이터)들을 하나의 그룹으로 묶는 것임

데이터들이 어떤 클러스터에 할당되어야 하는 지에 대한 정보(라벨링, y)가 없는 비지도 학습 알고리즘에 해당됨

각 개체들의 유사도(거리)정보를 기반으로 작동함



군집분석 및 시각화



군집분석(Clustering)

장점	단점
쉬운 계산 ➡ 다른 군집화 알고리즘에 비해 복잡도가 낮음	노이즈에 매우 민감
구현이 쉽고 다양한 언어와 플랫폼에서 제공되는 알고리즘	몇 가지 상황에서는 최적의 군집 구조를 찾기 어려움

군집분석 및 시각화



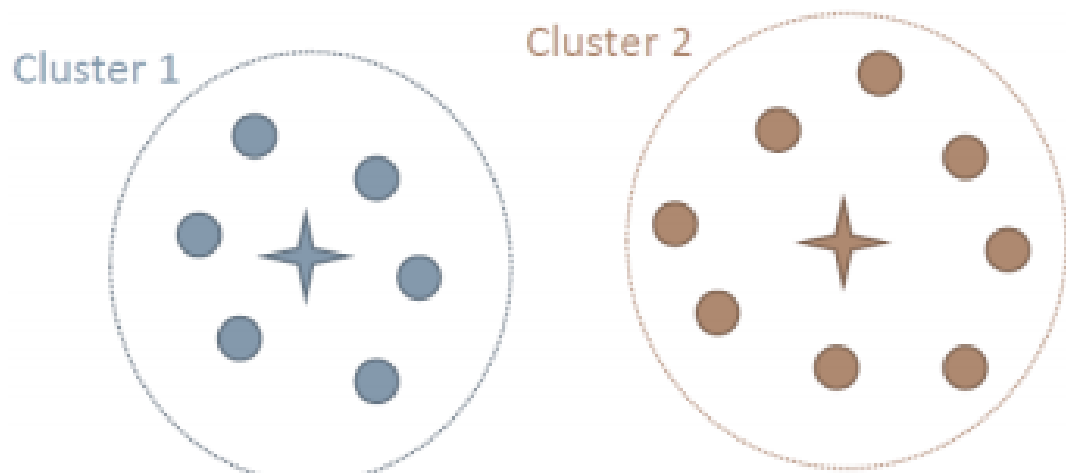
군집분석(Clustering)

◆ k-means Clustering

01

분할 군집화 방법

- 하나의 포인트는 반드시 하나의 군집에 소속



군집분석 및 시각화



군집분석(Clustering)

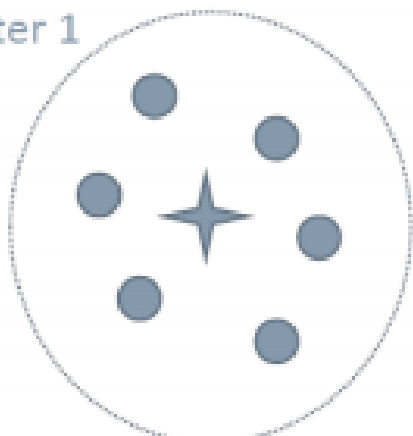
◆ k-means Clustering

02

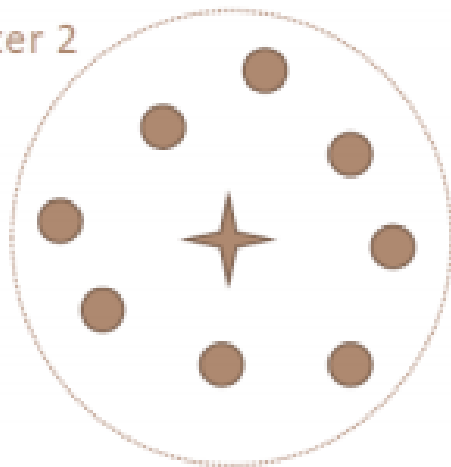
군집의 생성형태

- 각 군집은 하나의 중심을 갖고 있음
 - 이를 Centroid라고 부름
 - 각 포인트는 가장 가까운 Centroid에 해당하는 군집에 소속되

Cluster 1



Cluster 2



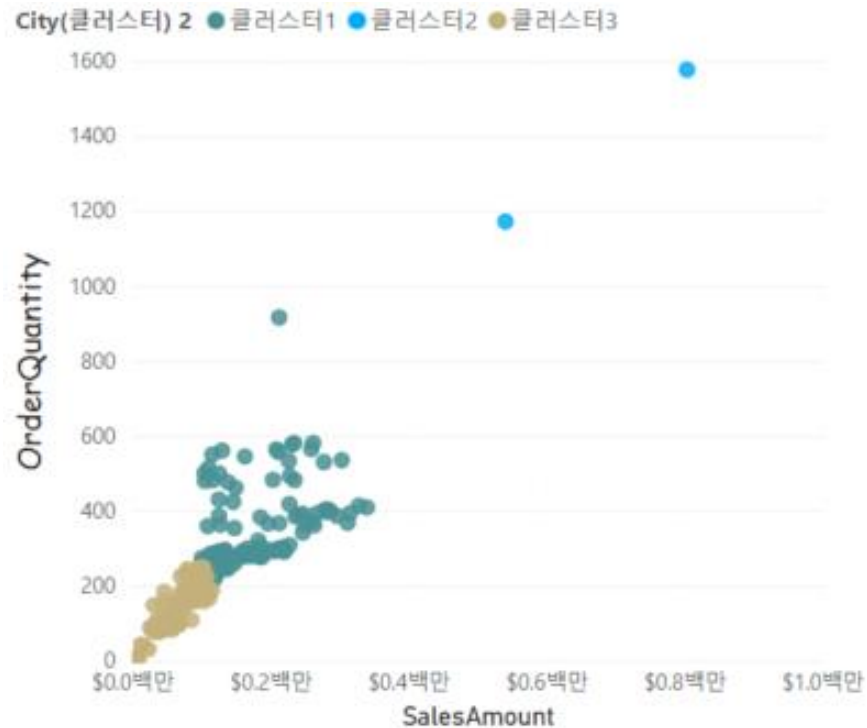
군집분석 및 시각화



군집분석(Clustering)

예

- 판매 지역을 매출액과 매출수량에 따라 3그룹으로 구분하라
- 자사상품을 매출합계와 총 판매 수량에 따라 4개의 그룹으로 분류하라



군집분석 및 시각화

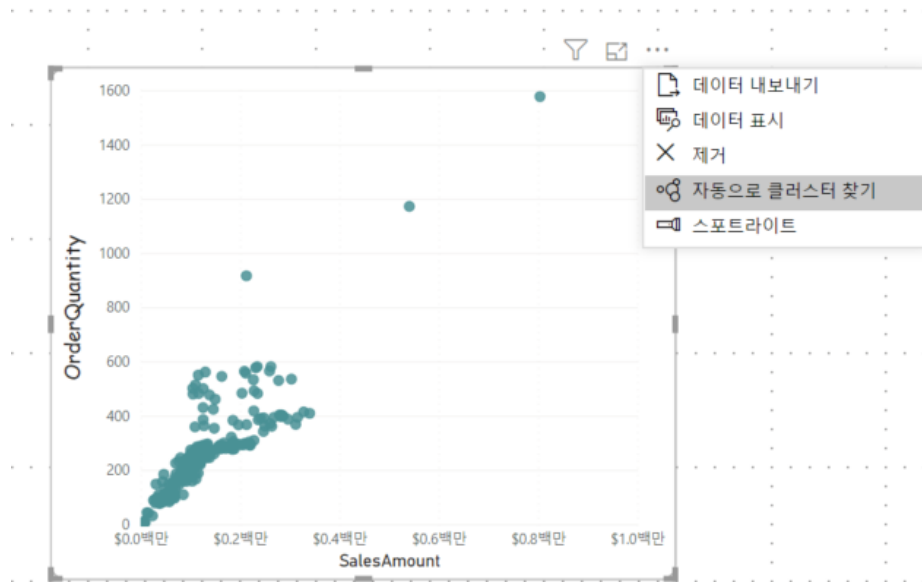


군집분석(Clustering)

예

■ 주의할 점

- 추가옵션에서 자동으로 클러스터 찾기 선택
- 범례 부분은 비워 두어야 함



자세히

City

범례

여기에 데이터 필드 추가

X축

SalesAmount

Y축

OrderQuantity

크기

여기에 데이터 필드 추가

군집분석 및 시각화



군집분석(Clustering)

◆ 클러스터링 옵션

01

Data preprocessing

- Scale data
 - 클러스터링 대상 독립변수의 수치를 동일한 스케일로 표준화
- Apply PCA
 - Principal Component Analysis 적용하기
 - 2개 이상의 독립변수가 적용되는 경우 권장함

군집분석 및 시각화



군집분석(Clustering)

◆ 클러스터링 옵션

02

Clusters definition

- Number of clusters
 - 클러스터의 개수(auto, 2-12)
- Method
 - 표시되는 클러스터의 개수를 자동화 하는 방법
(Fast, Moderate, Slow)

군집분석 및 시각화



군집분석(Clustering)

◆ 클러스터링 옵션

03

Visual appearance

- Draw ellipse
 - 군집을 표현하는 타원 그리기
 - ➡ 클러스터를 모두 포함하는 타원은 아님
- Draw convex hull
 - 클러스터 그룹을 감싸기 위한 최소한의 부분 집합을 표시
 - ➡ 즉 군집을 구성하는 가장 바깥 원소들을 이어 그린 것

군집분석 및 시각화



군집분석(Clustering)

클러스터링 옵션

04

Maximum Data

- Maximum exported rows
- 최대 행 개수 지정
 - 예 1000, 10000, 50000, Unlimited

