

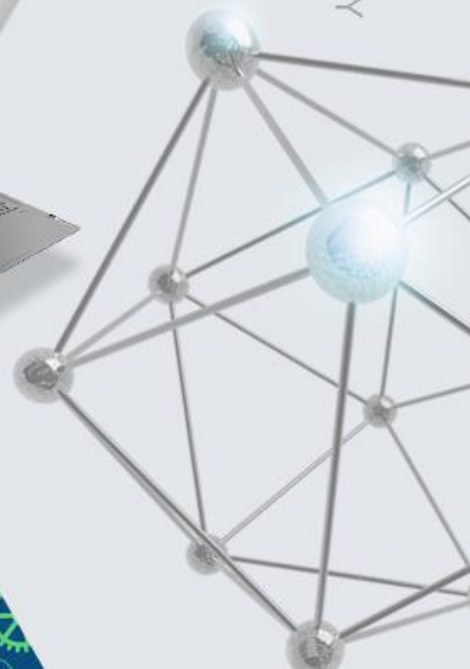


한국기술교육대학교
온라인평생교육원

The 4th Industrial Revolution is characterized by super connectivity and super intelligence, where various products and services are connected to the network, and artificial intelligence and information communication technologies are used in 3D printing, unmanned transportation, robotics. Of the world's most advanced technologies.

빅데이터 입문

예측적 빅데이터 기술분석 수행하기



The 4th Industrial Revolution is characterized by super connectivity and super intelligence, where various products and services are connected to the network, and artificial intelligence and information communication technologies are used in 3D printing, unmanned transportation, robotics. Of the world's most advanced technologies.





예측적 빅데이터 기술분석 수행하기



학/습/목/표

1. 예측적 빅데이터 분석 기법 중 회귀분석과 사례기반 추론을 파악할 수 있다.
2. 예측적 빅데이터 분석 기법 중 신경망분석과 의사결정나무기법을 파악할 수 있다.



학/습/내/용

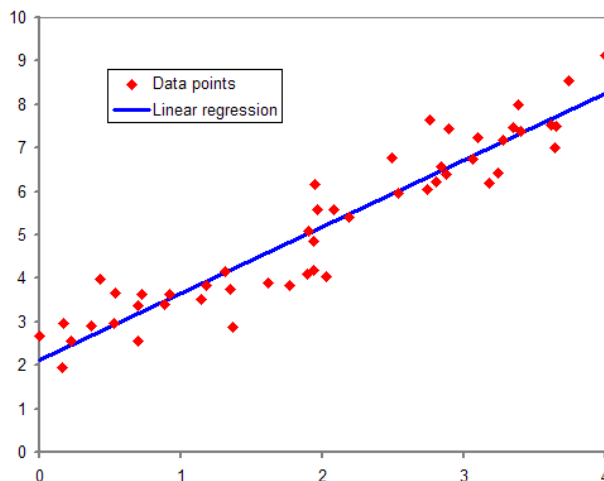
1. 회귀분석과 사례기반 추론
2. 신경망 분석과 의사결정나무

1. 회귀분석과 사례기반 추론

1) 회귀분석

(1) 예측모형

- 과거의 데이터를 기반으로 관측되지 않았던 변수에 미래의 값을 평가
- 통계적 지식 필요, 머신러닝 본격 적용
- 신제품 구매 고객, 재구매 고객, 불량품 확률 예측 가능
- 예측의 목적 설정이 필요함
- 예측값이 연속형이면 '예측'
- 예측값이 이산형이면 '분류'
- 예측모형을 만들 때 대표적으로 사용하는 방법: 회귀분석



- 대표적인 예측모델 방법
- 종속변수와 독립변수의 관계를 통계적으로 제시

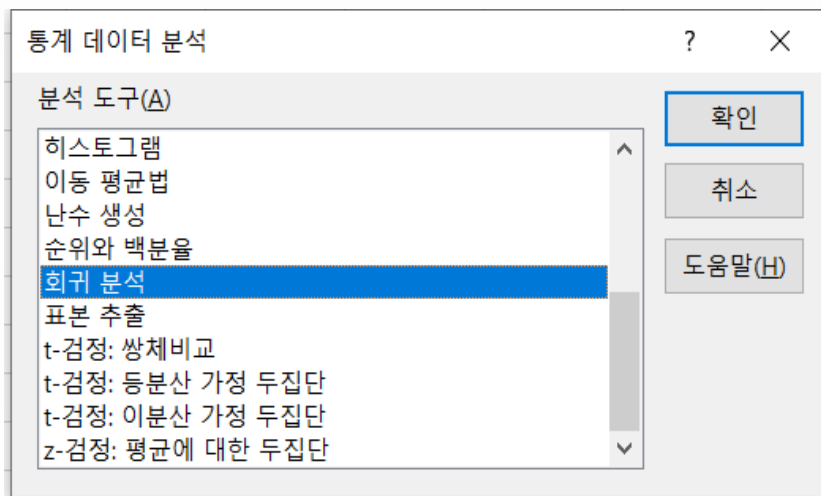
엑셀로도 쉽게 수행 가능함

1. 회귀분석과 사례기반 추론

1) 회귀분석

(2) 회귀분석

- 종속변수와 독립변수의 관계를 통계적으로 제시
- “어떤 변수가 다른 변수에 의하여 설명된다.”라고 가정하고 그 함수관계를 조사하는 통계적인 해석 기법



엑셀 상단메뉴 > 데이터 > 통계 데이터 분석 > 회귀분석

1. 회귀분석과 사례기반 추론

1) 회귀분석

(2) 회귀분석

- 회귀분석을 통해 Y축에 들어갈 값들, X축에 들어갈 값들을 선택
- Y축과 X축은 특정 기온이 얼마일 때 아이스크림에 판매량을 얼마가 될지를 우리가 회계분석 창에 대응
- 회귀분석 버튼을 클릭
- 회귀분석 통계량, 다중상관계수, 결정계수, 조정된 결정계수 값을 구할 수 있음

→ 회귀분석이라는 예측모형 완성

회귀 분석

입력

Y축 입력 범위(Y): ↑

X축 입력 범위(X): ↑

☒ 이분표(L) ☐ 상수에 0을 사용(Z)

☐ 신뢰 수준(F) %

출력 옵션

☐ 출력 범위(O):

☒ 새로운 워크시트(P):

☐ 새로운 통합 문서(W)

잔차

☐ 잔차(B) ☐ 잔차도(D)

☐ 표준 잔차(T) ☐ 선적할도(I)

정규 확률

☐ 정규 확률도(N)

예측모형을 만들 수 있음

	A	B	C	D	E	F	G	H	I
1	요약 출력								
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									



1. 회귀분석과 사례기반 추론

2) 사례기반 추론

(1) 사례기반추론 (CBR)의 정의

- 과거 사례의 결과를 바탕으로 새로운 사례의 결과를 예측하는 방법

(2) 사례기반추론 (CBR)의 4단계

- 검색: 현재 우리가 닦친 상황(사례)와 가장 유사한 적절한 사례를 검색하는 단계
- 재사용: 과거 사례의 해당 문제의 해결 방법을 접목하여 새로운 상황에 대한 적절한 해결방안을 적용하는 단계
- 수정: 새로운 해결 방법이 적합한지 테스트하여 적합하지 않은 부분은 수정하는 단계
- 유지: 성공적인 해결책을 발견한 후 이를 적용하고 데이터베이스에 새로운 사료로 저장하는 단계

(3) 사례기반추론 예시 - “2030년 새로운 팬데믹 발생”

- 검색: 세계보건기구 DB에서 검색
- 재사용: 대한민국 대처방법 재사용
- 수정: 적용이 어려운 부분은 테스트하여 수정
- 유지: 새로운 해결책으로 저장하여 유지



1. 회귀분석과 사례기반 추론

2) 사례기반 추론

(4) 사례기반추론 (CBR)의 장점

- 구조가 간단하고 결과의 예가 쉬움
 - 새로운 사례를 통한 학습이 가능
 - 적은 정보로 복잡한 문제를 해결, 의사결정을 지원
- 단순하고 쉽게 활용 가능

(5) 사례기반추론 (CBR)의 단점

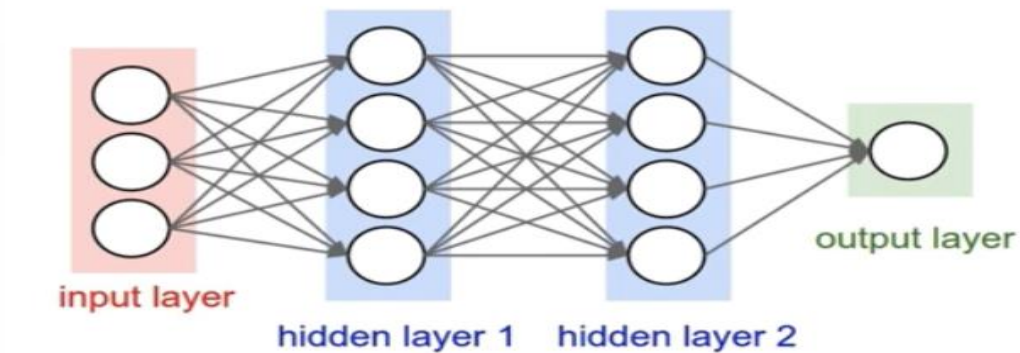
- 정확도가 많이 떨어짐
- 사례저장을 위한 데이터베이스 공간 많이 필요
- 일반화와 학습과정을 통한 많은 시간 소요

2. 신경망 분석과 의사결정나무

1) 신경망 분석 (Neural network)

(1) 신경망 분석이란?

- 인간 뇌의 뉴런 작동원리를 모방하는 방법
- 복수의 층(Layer)과 각 층 안에 노드(Node)가 연결되어 있는 구조



- 데이터 내부의 패턴이나 구조를 인지하는데 필요한 입력층, 은닉층, 출력층을 활용
- 처리요소에서 처리 후, 가중치를 결정
- 예측 정확도를 위하여 반복적으로 가중치를 수정하여 정확성을 높임

2. 신경망 분석과 의사결정나무

1) 신경망 분석 (Neural network)

(2) 네 가지 주요 특징

- 비선형기법으로 예측력이 높고, 통계적 가정이 필요 없음
- 스몰 데이터, 불완전 데이터가 많은 경우에도 효과적인 처리가능
- 모델학습에 과다한 시간소요, 결과 해석 난해
- 전체 최적해보다는 부분 최적해가 산출될 수 있음

(3) 대표적 활용 분야

- 영상인식, 화상인식, 음성인식 분야
- 주가예측, 부도예측, 신용평가
- 로봇제어, 무인 자동차
 - Nvidia와 같은 기업들의 GPU 기술 발전
 - GPU 기술을 활용한 딥러닝으로의 발전

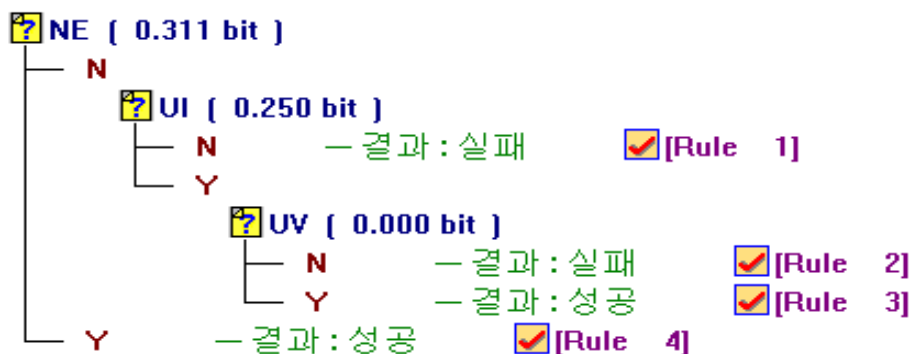
2. 신경망 분석과 의사결정나무

2) 의사결정나무

(1) 의사결정나무란?

- 특정 대안이 선택될 것인가라는 것과 발생할 수 있는 불확실한 상황 중에서 어떤 것이 실현되는지에 대하여 여러 결과가 생긴다는 상황을 나뭇가지와 같은 모양으로 도식화 한 것

Total Entropy : 1.000 bit



2. 신경망 분석과 의사결정나무



2) 의사결정나무

(2) 분석 4단계

- 분석목적에 부합하는 분리기준에 근거한 의사결정나무 산출
- 분류오류를 높이거나 부적절한 추론규칙 가지 제거
- 타당성 평가 : 정성분석 등을 통한 의사결정나무 평가
- 해석 : 해석과 예측모형 설정

(3) 알고리즘의 종류

- ID3 알고리즘
- CART 알고리즘
- CHAID 알고리즘

2. 신경망 분석과 의사결정나무



2) 의사결정나무

(4) 장점

- 변수 선정 및 선택이 용이
- 해석이 용이
- 지식추출의 용이

(5) 단점

- 예측 오류가 크다
- 비 안정성



1. 회귀분석과 사례기반 추론

1) 회귀분석

- 어떤 변수가 다른 변수에 의하여 설명된다고 보고 그 함수관계를 조사하는 통계적인 기법 중 하나

2) 사례 기반 추론

- 과거에 문제를 해결하는데 사용하였던 사례나 경험에 비추어 현재의 새로운 문제해결책을 제시하는 방법



2. 신경망 분석과 의사결정나무

1) 신경망 분석

- 인간 뇌의 뉴런 작동원리를 모방하는 방법으로 복수의 층(Layer)와 각 층 안에 노드(Node)가 연결되어 있는 구조

2) 의사결정나무

- 특정 대안이 선택될 것인가라는 것과 발생할 수 있는 불확실한 상황 중에서 어떤 것이 실현되는지에 대하여 여러 결과가 생긴다는 상황을 나무가지와 같은 모양으로 도식화한 것

