

자율시스템으로써의 인공지능

승실대학교 베어드교양대학
 서유희 교수
 yhsuh@ssu.ac.kr

목차

- ▶ 자율시스템의 개념
- ▶ 딥러닝의 원리
- ▶ 자율시스템과 블랙박스 모델

지난시간 배운 내용

주	주제	온라인	오프라인
1	인공지능의 과거 현재와 미래	1. 강의 및 교과목 소개(공통, 핵심만) 2. 인공지능의 과거와 현재 3. 인공지능의 미래와 다양한 시선 4. 인공지능 개발환경 구축과 사용법(Anaconda/Colab)	1. 강의 및 교과목 소개(분반별, 자세히) 2. 다양한 인공지능 기술 경험하기 (자연어처리, 시각, 음성) 3. 인공지능 챗봇 만들기(IBM 왓슨 어시스턴트)
2	공공데이터를 이용한 사회문제 발견과 해결책 모색	1. 빅데이터의 정의와 가치 2. 공공데이터 수집하기 3. 공공데이터로부터 새로운 인사이트 발견하기 - 행정구역별 인구 데이터와 공공의료기관 현황 데이터 분석	1. 서울시 CCTV설치 현황 분석하기 2. 서울시 범죄발생 현황 분석하기
3	인공지능의 개요 및 머신러닝을 이용한 예측	1. 인공지능의 정의와 분류 2. 인공지능 학습방법 이해하기 3. 인공지능 알고리즘 소개	1. 머신러닝을 이용한 이미지 식별(구글 티처블 머신) 2. 머신러닝을 이용한 보스턴 집값 예측
4	인공지능과 데이터 윤리	1. 데이터의 불완전성과 결함에 따른 예측 오류와 차별 2. 데이터 편향성이 예측에 미치는 영향 (구글 티처블 머신) 3. 지도학습(SVM)을 이용한 타이타닉호 생존자 예측	1. 타이타닉호 생존자 예측 - 데이터 편향성이 예측에 미치는 영향 - 데이터 왜곡에 따른 예측 결과 비교
5	인공지능과 알고리즘 윤리	1. 알고리즘과 모델링의 개요 2. 알고리즘 기반 의사결정 시스템의 한계 3. 윤리가 필요한 인공지능 4. 오렌지3 설치 및 사용법	1. 오렌지3를 이용한 알고리즘에 따른 예측 결과 비교 - 보스턴 집값 예측 - 폐암환자 생존 여부 예측
6	인공지능에 대한 윤리적 쟁점과 다양한 이슈	1. 자율시스템으로써의 인공지능과 딥러닝 2. 인공지능 안전성과 윤리 3. 인공지능의 윤리적 쟁점 (자율주행자동차, AI로봇, 트랜스휴먼, 프라이버시 문제)	1. 비윤리적 데이터 생성과 수집 - 웹 스크래핑(크롤링)을 이용한 데이터 수집
7		기말고사	

학습 목표

- ▶ 자율시스템이란 무엇인지 설명할 수 있다.
- ▶ 오늘날 자율시스템을 구현하기 위한 딥러닝 기술을 이해한다.
- ▶ 자율시스템과 블랙박스 모델의 문제점을 설명할 수 있다.

인공지능의 분류

5

3 stages of AI



좁은 인공지능 (Narrow AI)

- 스스로 사고해 문제를 해결할 수 있는 능력이 없는 컴퓨터 기반 AI
- 특정분야에 국한된 인공지능
- 특정 영역에서 인간보다 나은 성과
- AlphaGo, Watson, Siri



일반 인공지능 (General AI)

- 스스로 사고해 문제를 해결할 수 있는 능력을 가진 컴퓨터 기반 AI
- 인간 수준의 능력을 가진 인공지능
- 모든 분야에 적용될 수 있는 인공지능
- 자각력과 독립성을 갖춘



수퍼 인공지능 (Super AI)

- 인간보다 100~1000배 (IQ 1만~10만) 뛰어난 지능을 가진 AI
- 효율, 자기보존, 자원획득, 창의성 같은 원초적 욕구를 기반으로 끊임없이 자기 발전

이미지 출처 : <https://www.datakeen.co/en/what-is-artificial-intelligence/>

인공지능의 용어간 관계

6

Artificial Intelligence Simulates human intelligence

Narrow AI
Specific tasks

Machine Learning Learns on past data

Deep Learning A particular learning structure - deep neural networks

General AI
~ Human
Super AI
>> Human

Now

Future (?)

- 인공지능(Artificial Intelligence)
- 인간과 비슷하게 사고하는 컴퓨터의 지능을 구현하는 포괄적 개념

- 머신러닝(Artificial Intelligence)
- 데이터를 입력해 컴퓨터를 학습시키거나 스스로 배우게 해 인공지능 성능을 향상시키는 방법

- 딥러닝(Deep Learning)
- 머신러닝의 한 분야로 신경망(Neural Network)을 통해서 학습하는 방법

이미지 출처 : <https://www.datakeen.co/en/what-is-artificial-intelligence/>

자율시스템

7

▶ 스스로 인식, 계획, 학습, 진단, 제어, 중재, 협업 등 지능적 기능들을 가지는 시스템

▶ 자율성을 정량적으로 측정, 평가하기 위한 자율성 수준

▶ 미국의 NIST 프로젝트에서 개발한 ALFUS(Autonomy Level for Unmanned System)

- ▶ 임무복잡성, 환경복잡성, 인간독립성 세가지 측면에 따라 0~10 수준으로 분류
- ▶ 자율성 수준 0: 원격 조종 단계
- ▶ 자율성 수준 1~4 : 정해진 프로그램에 의해 작동되는 자동 조종 단계
- ▶ 자율성 수준 5~9 : 의사결정기능을 갖춘 자율 조종 단계

▶ 미국 자동차공학회 SAE(Society of Automotive Engineers) 자율주행자동차 기술 등급

- ▶ 자율성 수준에 따라 6단계로 분류
 - ▶ 레벨0 : 비자율 주행
 - ▶ 레벨1 : 운전자보조, 레벨2 : 부분자율주행
 - ▶ 레벨3 : 조건부 자율주행
 - ▶ 레벨4 : 고도자율주행
 - ▶ 레벨5 : 완전자율주행

딥러닝 (Deep Learning)

8

▶ 딥러닝

▶ 현재 자율시스템(스스로 의사 결정)을 구현하기 위한 주요 기술

▶ 머신러닝의 다양한 알고리즘의 한 종류인 인공신경망 알고리즘이 발전된 형태의 기계 학습 방법

- ▶ 방대한 양의 학습 데이터와 아주 많은 연산을 수행해야 함으로 높은 성능 하드웨어가 필요
- ▶ 인간의 뇌가 학습하는 방법을 수학적으로 모델링
- ▶ 머신러닝에 비해 학습 방법이 오래걸림

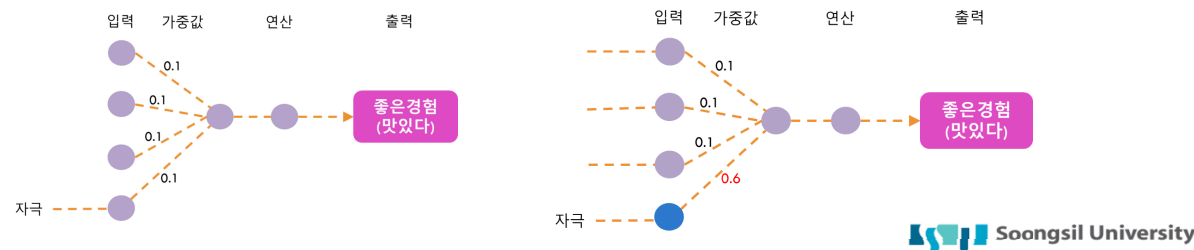
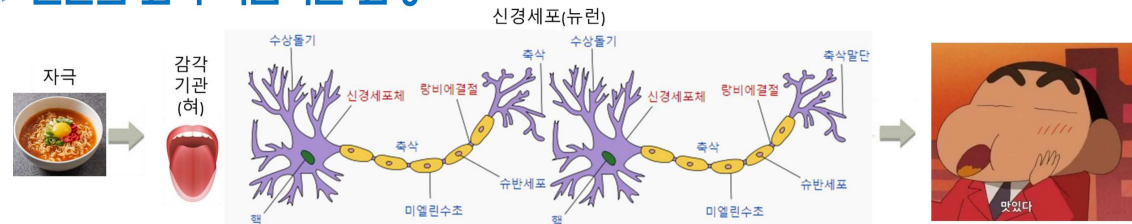
▶ 인간의 개입을 통해 학습하는 머신러닝(모델 기반 기계학습)에서 기계 스스로 학습하는 능력이 강화된 인공지능(스스로 좋은 모델을 찾아가는 기계학습)

▶ 인간의 개입이 최소화되어 설계된 모델로 스스로 복잡한 심층학습을 통해 내부 신경망의 연결이 증가되고 복잡해짐

인공신경망(Artificial Neural Networks)

9

▶ 인간의 뇌가 학습하는 과정

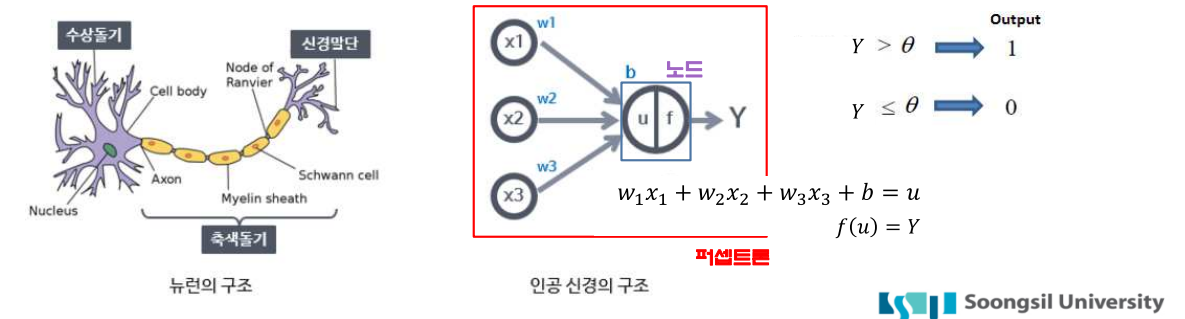


인공신경망(Artificial Neural Networks)

10

▶ 인공신경망 구조

- ▶ input(x)이 들어오면 가중치(w)가 곱해짐
- ▶ 그 값들의 합(Y)이 어느 임계값보다 크면 1을 다음 노드(뉴런)의 입력으로 보냄 (0은 출력이 없는 상태와 같음)
- ▶ 0과 1을 판단하는 함수를 **활성화 함수(activation function)**이라고 함

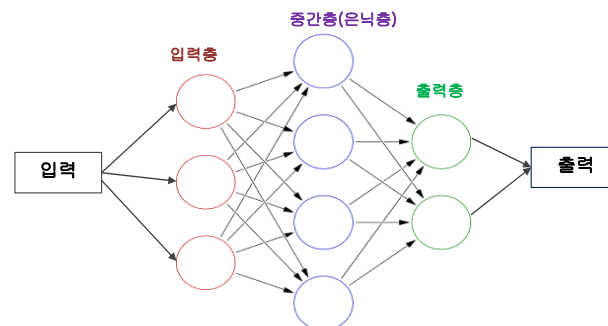


인공신경망(Artificial Neural Networks)

11

▶ 다층 인공 신경망의 구조

- ▶ 노드가 서로 연결된 구조의 신경망
- ▶ 입력층에 학습시키고 싶은 데이터를 입력
- ▶ 데이터들이 입력층, 중간층(은닉층), 출력층을 지나며 처리되고 최종결과 출력

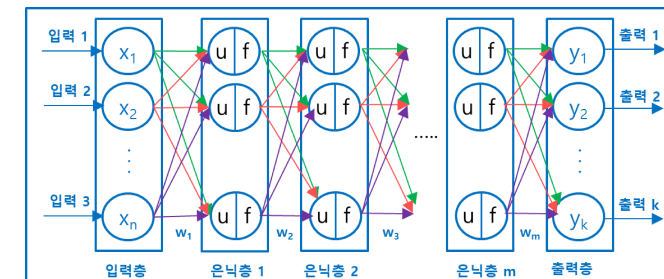


딥러닝

12

▶ 딥러닝의 학습

- ▶ **딥러닝의 학습**이란 다층 신경망 구조에서 출력층에서의 오차가 최소가 되도록 각 노드(뉴런)을 연결하는 **접속 가중치와 편향 값을 학습과정**
- ▶ 노드가 서로 연결되어 있는 신경망 구조에서 1개 이상의 은닉층을 두어(은닉층을 깊게 할수록) 학습하면 정확도가 높은 결과를 얻을 수 있다하여 사용된 이름

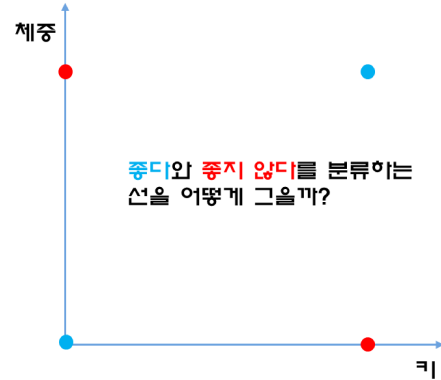


XOR 문제

13

- 키와 체중의 데이터를 이용해 건강상태를 분류하는 예
 - 건강한 그룹(●)과 건강하지 않은 그룹(○)으로 분류

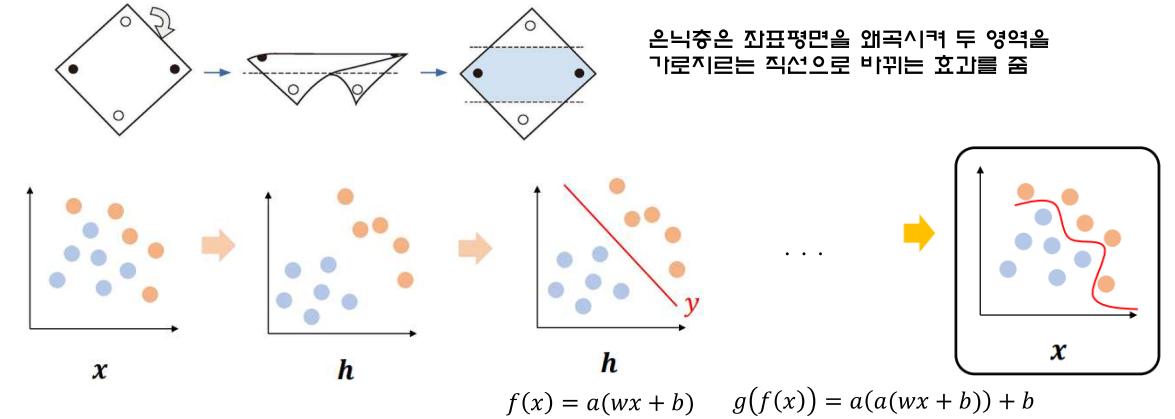
키	체중	건강
작다(0)	많다(0)	● 좋지않다(0)
크다(1)	많다(0)	○ 좋다(1)
작다(0)	적다(1)	○ 좋다(1)
크다(1)	적다(1)	● 좋지않다(0)



은닉층의 역할

14

- 은닉층은 선형으로 분리되지 않는 클래스를 선형으로 분리되도록 만들어 줌



자율시스템과 블랙박스 모델

15

- 자율시스템에는 최소한의 제한된 규칙만 있을 뿐 정해진 입출력 법칙이 없음
- 모델의 설계자조차 내부 학습 과정을 추적하기 어렵기 때문에(블랙박스 모델) 최종 출력의 이유도 알기 어려움
- 내부 작동과정을 알 수 없는데 인공지능의 판단을 신뢰할 수 있을까?
 - 예) 의료 인공지능, 인공지능 변호사, 불량품 검출, 업무역량평가, 무인 자동차 등
 - 인공지능의 블랙박스의 투명성, 설명가능성에 대한 요구
- 인공지능의 잘못된 판단에 대한 책임은 누가 질 수 있을까?
 - 인공지능의 불투명성으로 인해 잘못된 판단을 했을 때 사회적 대처와 처리가 어려움
 - AI의사의 진단오류, AI판사의 잘못된 판결 등, 무인자동차의 교통사고 등
 - 인공지능의 자율의사결정에 대한 책임성 확보에 대한 요구

다음시간에 배울 내용

16

주	주제	온라인	오프라인
1	인공지능의 과거 현재와 미래	1. 강의 및 교과목 소개(공통, 핵심만) 2. 인공지능의 과거와 현재 3. 인공지능의 미래와 다양한 시선 4. 인공지능 개발환경 구축과 사용법(Anaconda/Colab)	1. 강의 및 교과목 소개(분반별, 자세히) 2. 다양한 인공지능 기술 경험하기 (자연어처리, 시각, 음성) 3. 인공지능 챗봇 만들기(IBM 왓슨 어시스턴트)
2	공공데이터를 이용한 사회문제 발견과 해결책 모색	1. 빅데이터의 정의와 가치 2. 공공데이터 수집하기 3. 공공데이터로부터 새로운 인사이트 발견하기 - 행정구역별 인구 데이터와 공공의료기관 현황 데이터 분석	1. 서울시 CCTV설치 현황 분석하기 2. 서울시 범죄발생 현황 분석하기
3	인공지능의 개요 및 머신러닝을 이용한 예측	1. 인공지능의 정의와 분류 2. 인공지능 학습방법 이해하기 3. 인공지능 알고리즘 소개	1. 머신러닝을 이용한 이미지 식별(구글 티쳐블 머신) 2. 머신러닝을 이용한 보스톤 집값 예측
4	인공지능과 데이터 윤리	1. 데이터의 불완전성과 결함에 따른 예측 오류와 차별 2. 데이터 편향성이 예측에 미치는 영향 (구글 티쳐블 머신) 3. 지도학습(SVM)을 이용한 타이타닉호 생존자 예측	1. 타이타닉호 생존자 예측 - 데이터 편향성이 예측에 미치는 영향 - 데이터 왜곡에 따른 예측 결과 비교
5	인공지능과 알고리즘 윤리	1. 알고리즘과 모델링의 개요 2. 알고리즘 기반 의사결정 시스템의 한계 3. 윤리가 필요한 인공지능 4. 오픈지3 설치 및 사용법	1. 오픈지3를 이용한 알고리즘에 따른 예측 결과 비교 - 보스톤 집값 예측 - 폐암환자 생존 여부 예측
6	인공지능에 대한 윤리적 쟁점과 다양한 이슈	1. 자율시스템에서의 인공지능과 딜러닝 2. 인공지능 안전성과 윤리 3. 인공지능의 윤리적 쟁점 (자율주행자동차, 시로봇, 트랜스휴먼, 프라이버시 문제)	1. 비윤리적 데이터 생성과 수집 - 웹 스크래핑(크롤링)을 이용한 데이터 수집
7		기말고사	