

AI in Medicine

2019. 11. 20

김 화 종 (강원대학교 IT대학)

hjkim3@gmail.com

Contents

- ▶ What is AI
- ▶ Machine Learning
- ▶ AI in Medicine
- ▶ What to Do

Phase Changes

- ▶ 현미경, 전자장비에 이어 AI가 완전히 새로운 변화를 가져올 것



Hospital

- ▶ Checklist - Atul Gawandi
- ▶ 조직 문화를 바꾸는 것은 매우 어렵다



Surgical Safety Checklist			World Health Organization	Patient Safety A World Alliance for Safer Health Care
Before induction of anaesthesia (with at least nurse and anaesthetist)	Before skin incision (with nurse, anaesthetist and surgeon)	Before patient leaves operating room (with nurse, anaesthetist and surgeon)		
<p>Has the patient confirmed his/her identity, site, procedure, and consent?</p> <p><input type="checkbox"/> Yes</p>	<p><input type="checkbox"/> Confirm all team members have introduced themselves by name and role.</p> <p><input type="checkbox"/> Confirm the patient's name, procedure, and where the incision will be made.</p>	<p>Nurse Verbally Confirms:</p> <p><input type="checkbox"/> The name of the procedure</p> <p><input type="checkbox"/> Completion of instrument, sponge and needle counts</p> <p><input type="checkbox"/> Specimen labelling (read specimen labels aloud, including patient name)</p> <p><input type="checkbox"/> Whether there are any equipment problems to be addressed</p>		
<p>Is the site marked?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> Not applicable</p>	<p>Has antibiotic prophylaxis been given within the last 60 minutes?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> Not applicable</p>	<p>To Surgeon, Anaesthetist and Nurse:</p> <p><input type="checkbox"/> What are the key concerns for recovery and management of this patient?</p>		
<p>Is the anaesthesia machine and medication check complete?</p> <p><input type="checkbox"/> Yes</p>	<p>Anticipated Critical Events</p> <p>To Surgeon:</p> <p><input type="checkbox"/> What are the critical or non-routine steps?</p> <p><input type="checkbox"/> How long will the case take?</p> <p><input type="checkbox"/> What is the anticipated blood loss?</p> <p>To Anaesthetist:</p> <p><input type="checkbox"/> Are there any patient-specific concerns?</p> <p>To Nursing Team:</p> <p><input type="checkbox"/> Has sterility (including indicator results) been confirmed?</p> <p><input type="checkbox"/> Are there equipment issues or any concerns?</p>			
<p>Is the pulse oximeter on the patient and functioning?</p> <p><input type="checkbox"/> Yes</p>	<p>Is essential imaging displayed?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> Not applicable</p>			
<p>Does the patient have a:</p> <p>Known allergy?</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Yes</p>				
<p>Difficult airway or aspiration risk?</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Yes, and equipment/assistance available</p>				
<p>Risk of >500ml blood loss (7ml/kg in children)?</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Yes, and two IVs/central access and fluids planned</p>				

Digital Transformation

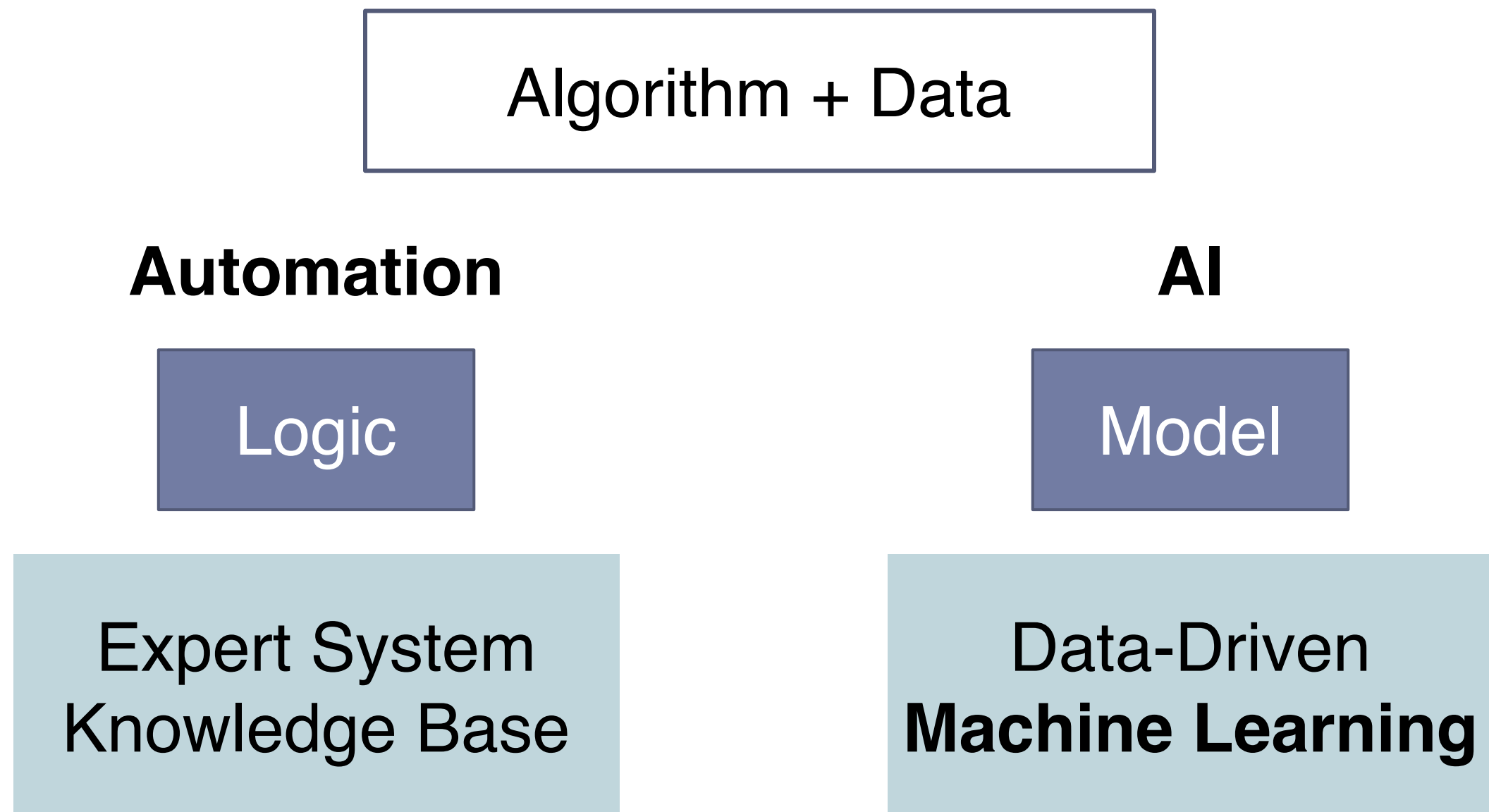
- ▶ 아날로그 음악은 디지털로 듣는다
- ▶ 돈의 사용은 대부분 디지털로 바뀌었다.
- ▶ 3D 프린팅의 핵심은 물체가 '디지털'로 표현되고 이동되는 것
- ▶ 건강정보 이용도 디지털로 바뀔 것



What is AI

Automation vs. AI

- ▶ 모두 알고리즘과 데이터를 사용한다. 그러나 구현 방법이 전혀 다르다

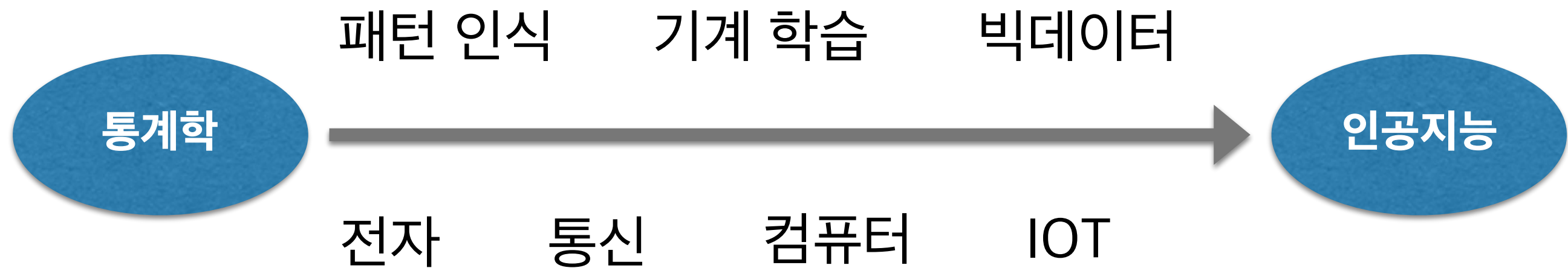


AI의 발전

- ▶ 영역별로 인간의 능력을 증가하고 있다.
 - ▶ 이미지 인식 – 보는 능력
 - ▶ 음성인식, 텍스트 인식 – 듣는 능력
 - ▶ 실시간 번역 – 말하는 능력
 - ▶ 이미지 캡션, 언어 모델링 – 쓰는 능력
 - ▶ 감성 능력까지 ?



AI 구현 기술



Terms

- ▶ 인공지능:
 - ▶ 지능이 있는 것처럼 컴퓨터가 **똑똑하게** 일을 처리하는 것 (구현 기술과 무관)
- ▶ 데이터마이닝:
 - ▶ 기 구축된 데이터베이스에서 **새로운** 지식을 얻는 것
- ▶ 비즈니스 인텔리전스:
 - ▶ 데이터 분석을 통해 새로운 **비즈니스** 전략을 얻는 것
- ▶ 통계분석:
 - ▶ **샘플 데이터**로부터 전체 데이터의 속성을 파악하는 것 (오차와 신뢰도 검증)
- ▶ 빅데이터 분석:
 - ▶ 대량의 데이터를 분석하여 일반적인(통계적인) 분석으로는 찾지 못하던 **새로운 가치**를 얻는 것
- ▶ 머신러닝:
 - ▶ 데이터로부터 새로운 지식을 얻는 **모델**을 만들고 학습으로 **모델의 성능을 개선**하는 것
- ▶ 딥러닝: 기계학습의 한 방법으로 최근의 인공지능의 **수준을 한 차원 높인** 기술

Big Data Analysis



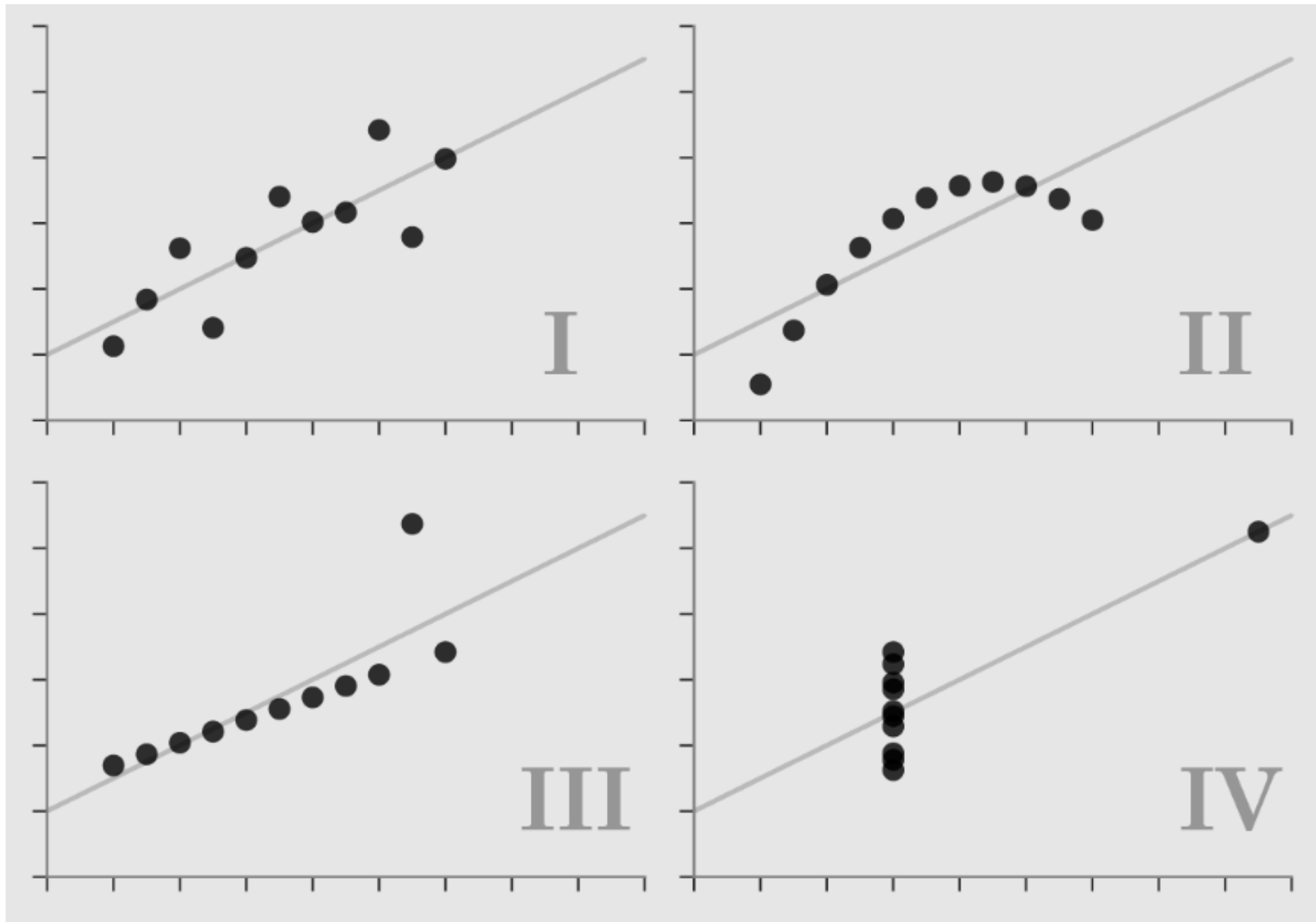
TARGET

- ▶ 상품 구매 이력을 보고 출산 예정일까지 예측한다



Correlation

- ▶ 아래는 모두 같은 상관계수를 갖는 분포이나 특성은 다르다



Machine Learning

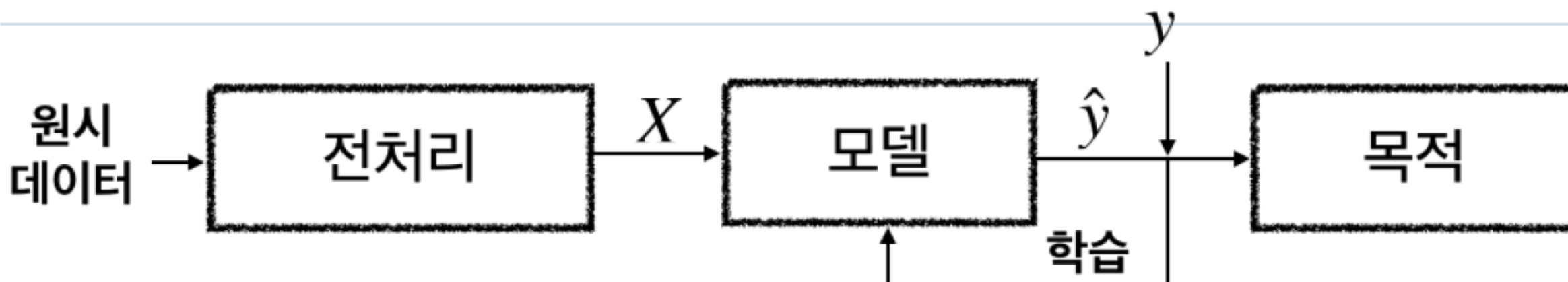
AI와 Machine Learning

- ▶ 인공지능을 구현하는 방법은 다양하다
 - ▶ Cognitive 컴퓨팅
 - ▶ 언어 기호학
 - ▶ 머신러닝 방식
- ▶ 머신러닝 기반의 AI가 2000년대 이후 급속히 발전
 - ▶ 머신러닝: 마치 사람이 많은 정보에 접하면서 학습하듯이 컴퓨터도 데이터를 보고 학습하는 방법
- ▶ 딥러닝: 신경망을 기반으로 하는 머신러닝 기술

머신러닝의 목적

- ▶ 설명(Descriptive analysis)
 - ▶ 통계적 특성 분석
 - ▶ 탐색적 분석 (EDA)
 - ▶ 클러스터링
- ▶ 예측(Predictive analysis)
 - ▶ 회귀 (regression)
 - ▶ 분류 (classification)
- ▶ 추천(Recommendation)
- ▶ 강화학습 (Reinforcement learning)

머신러닝 프로세스



결측치 처리
오류값 처리
스케일링
데이터 변환
· 카테고리 변환
· 로그, 역수 변환
특성공학
· 차원축소
· PCA

선형모델
로지스틱회귀
SVM
결정트리
랜덤포레스트
그라디언트부스트
kNN, Bayes
CNN
RNN

클러스터링
설명적 분석
· EDA, 시각화
회귀 예측
분류 예측
추천

데이터 전처리(preprocessing)

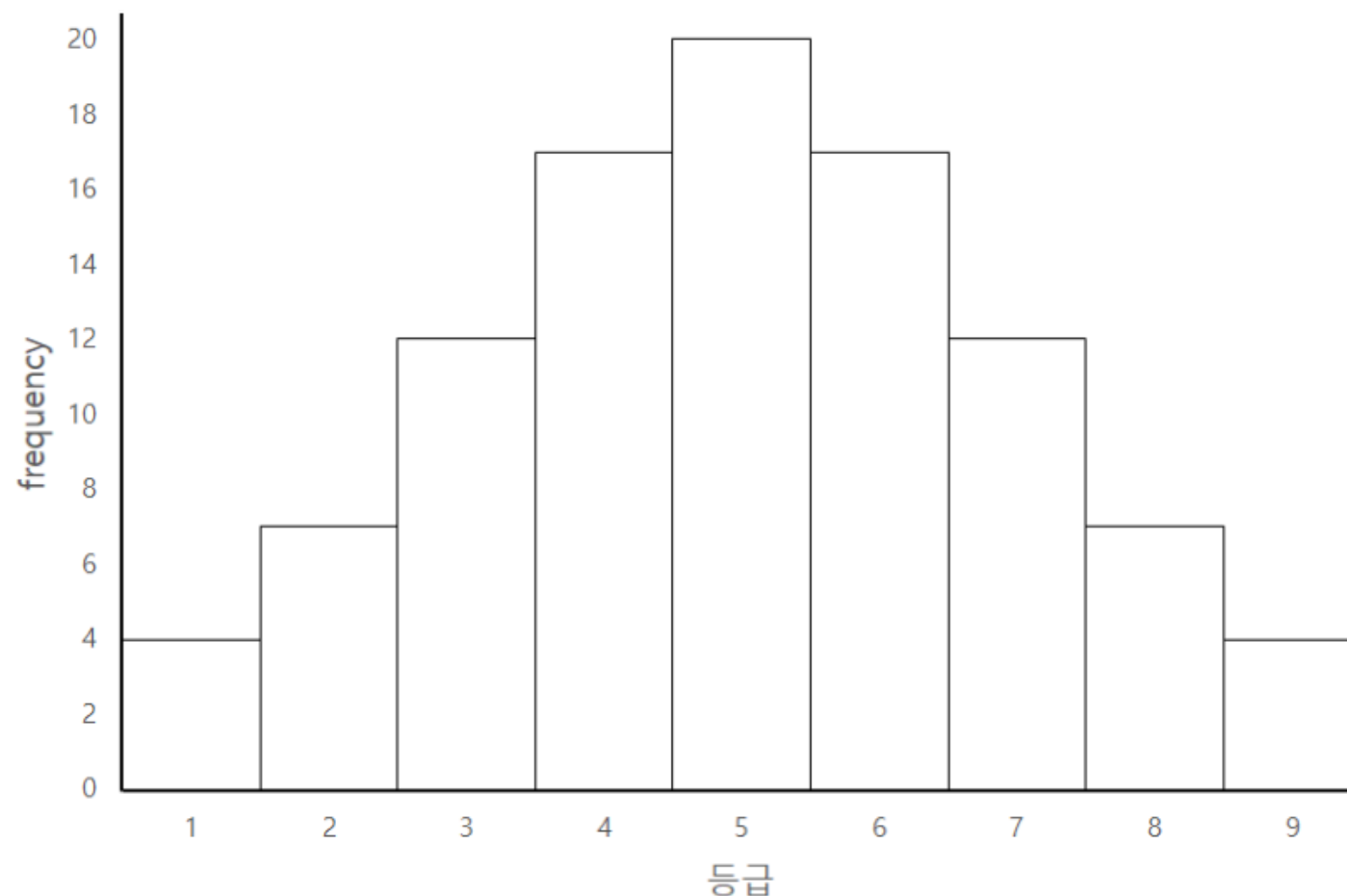
- ▶ 수집한 데이터를 분석하기 좋게 변환하는 작업
 - ▶ 데이터 클리닝 (결측치, 오류값 등의 처리)
 - ▶ 스케일링 (표준 스케일링, min-max 스케일링, Robust 스케일링)
 - ▶ 데이터변환 (카테고리 변환, 로그 변환, 역수 변환)
 - ▶ 특성 선택 (차원축소, PCA)
 - ▶ 클러스터링

탐색적 데이터 분석

- ▶ 데이터의 전체적인 특성을 파악
 - ▶ Exploratory Data Analysis: EDA
- ▶ 수집한 데이터가 분석에 적절한지 알아보는 과정
- ▶ 시각화 도구사용
 - ▶ 타임라인
 - ▶ 스캐터플롯
 - ▶ 박스플롯
 - ▶ 히스토그램
 - ▶ 바플롯

Categorical Encoding

- ▶ 나이 => 10대, 20대, 30대, 40대
- ▶ 연간 소득 => 고소득층, 중간층, 저소득층
- ▶ 내신 등급 분포 (등급 차이에 대한 효과가 같도록 정한다)
 - ▶ 입출력이 선형관계를 이루도록 한다

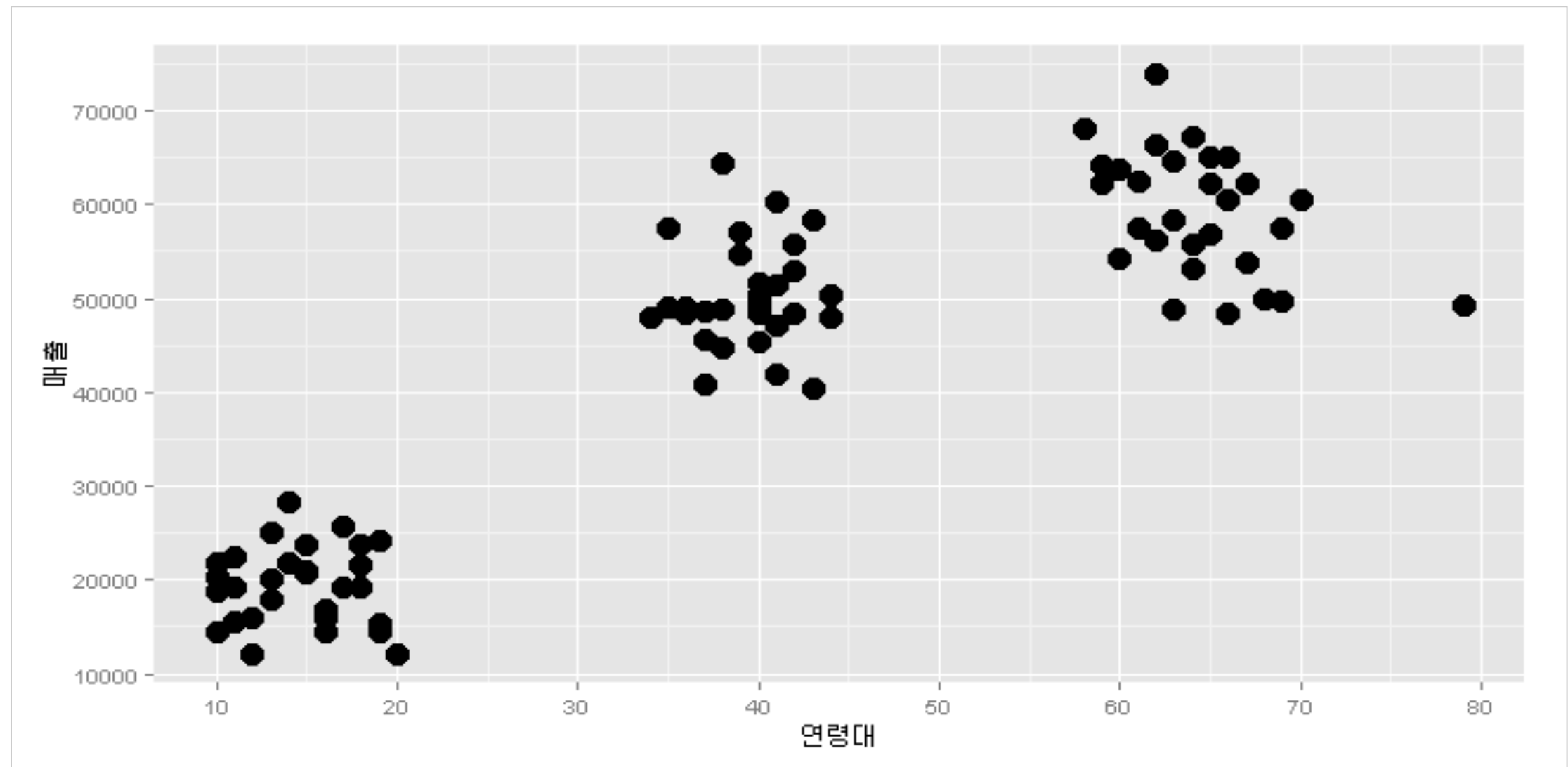


로그 변환

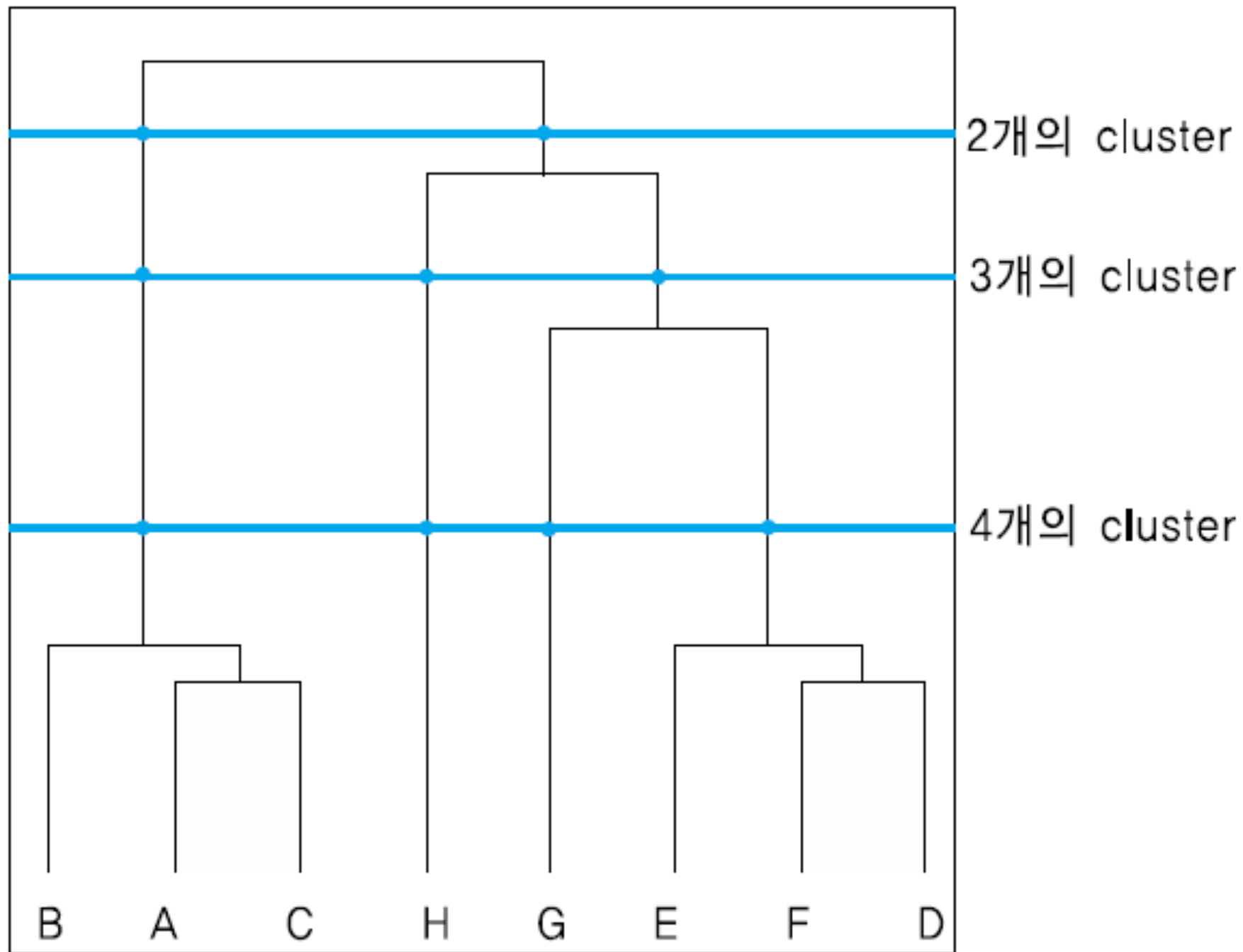
- ▶ 체감형 수치를 선형적으로 표현할 때 사용
 - ▶ 돈, 소리, 빛, 압력, 냄새 등 생물학적인 자극
- ▶ 같은 자극 변화를 느끼려면 현재 보유한 양이 많을수록 이에 비례한 더 강한 자극이 필요하다는 것
- ▶ 로그를 취한 이후의 값에 대해서 입출력의 관계가 선형적

클러스터링 (clustering)

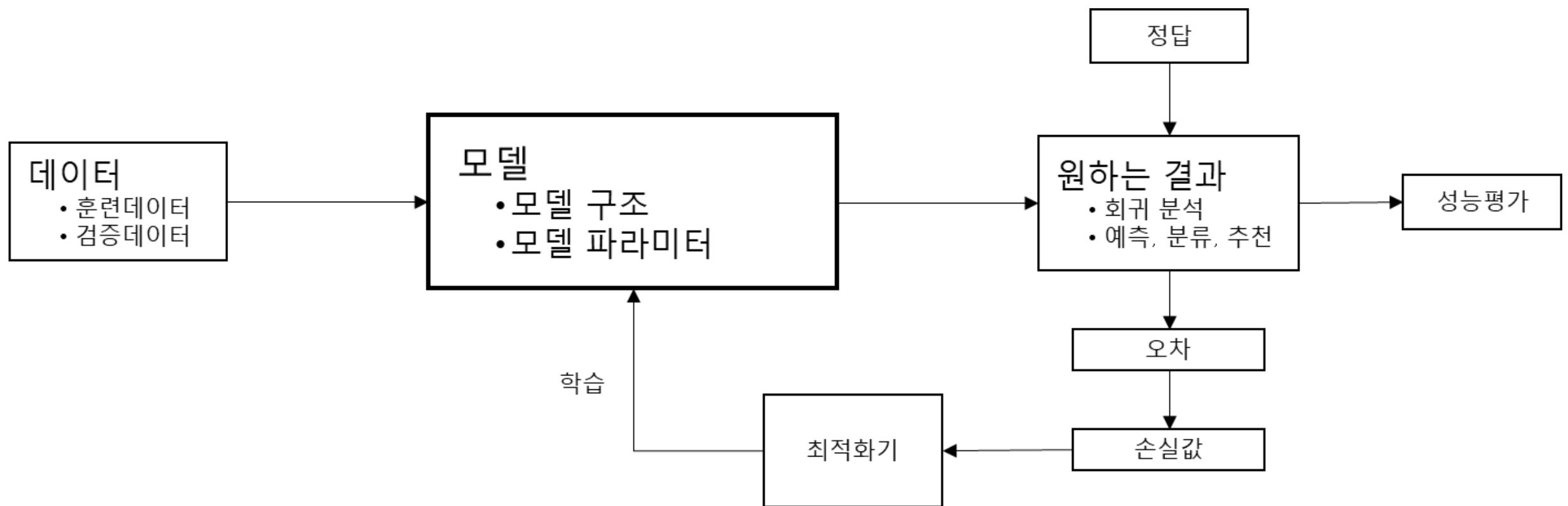
- ▶ 특성 공간에서 서로 가까운, 성격이 비슷한 항목들을 그룹으로 묶는 작업



덴드로그램(dendrogram)



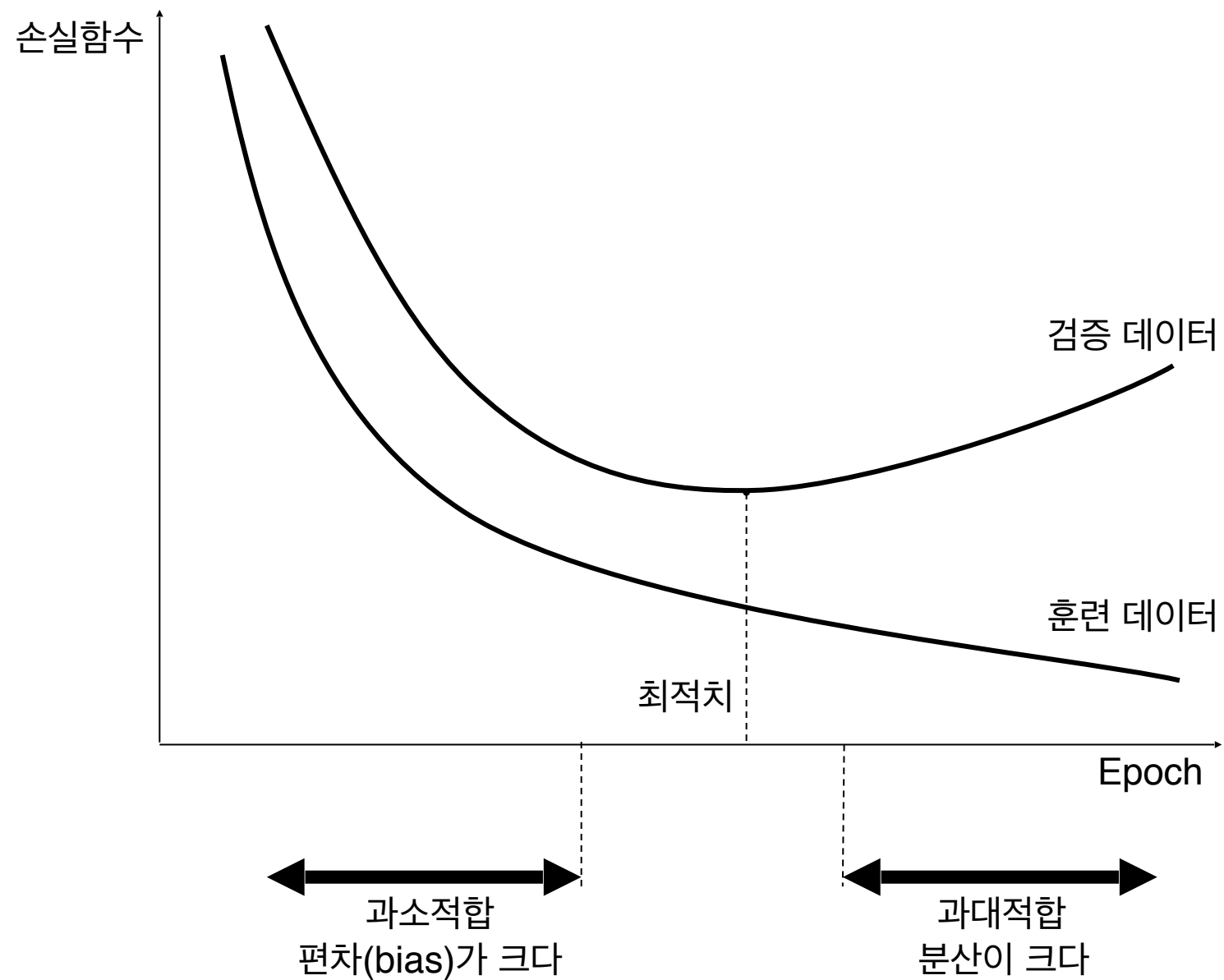
오차, 손실함수, 최적화, 파라미터



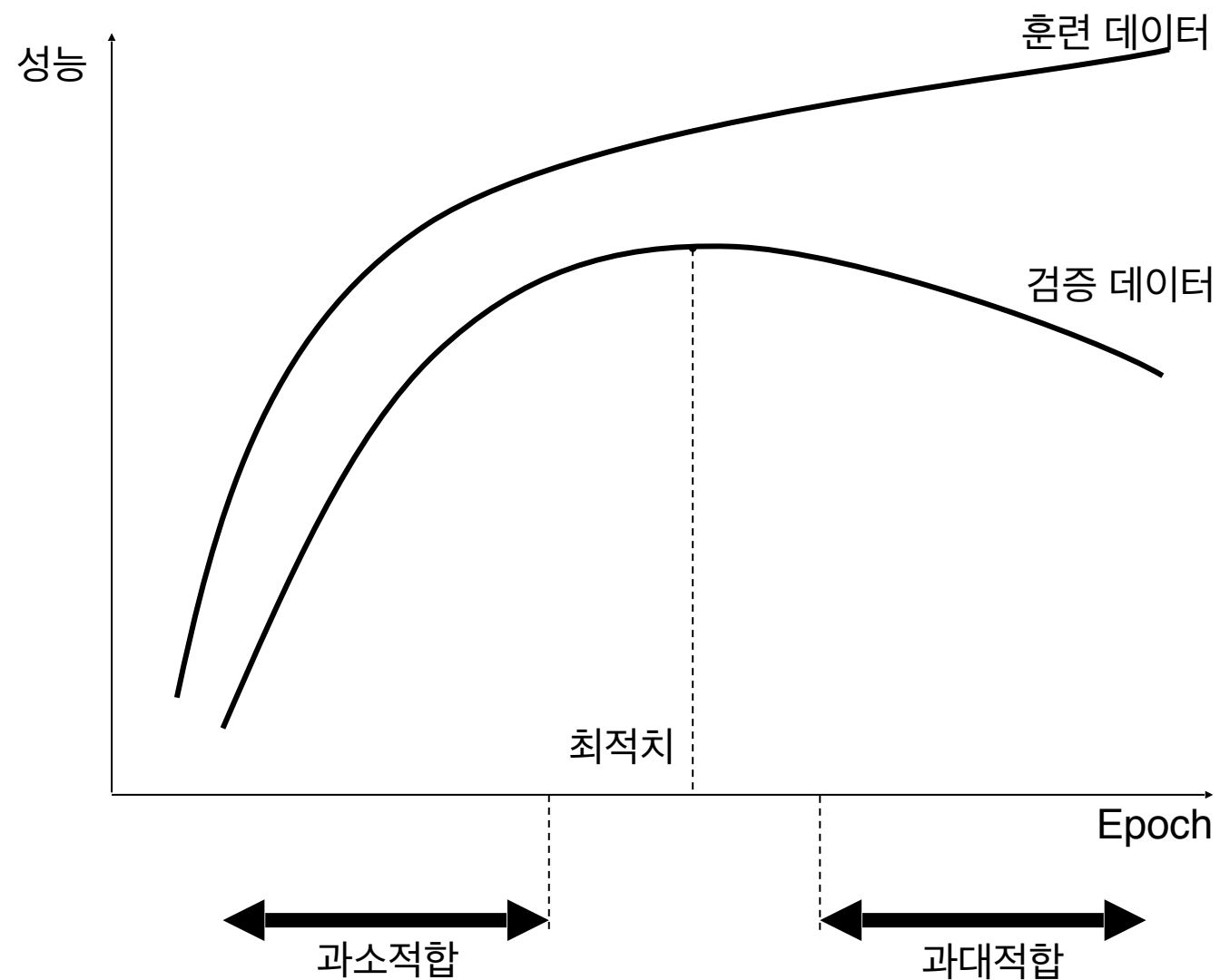
Loss Function vs. Performance

	회귀	분류
Loss Function	MAE, MSE, MSLE	Cross Entropy
Performance Measure	R Squared	Accuracy Precision Recall F-1 score ROC AUC

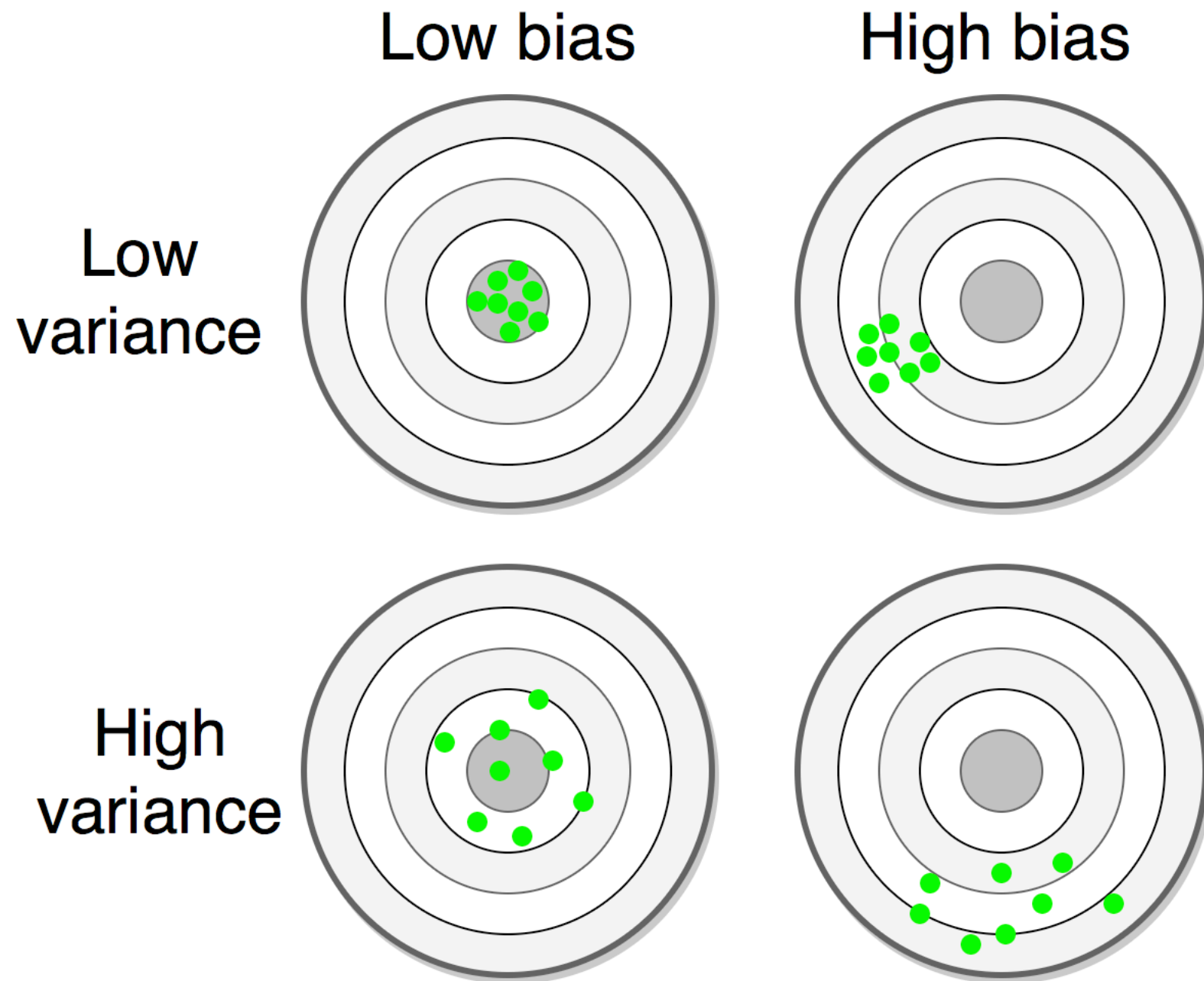
Overfitting vs. Underfitting



Overfitting vs. Underfitting

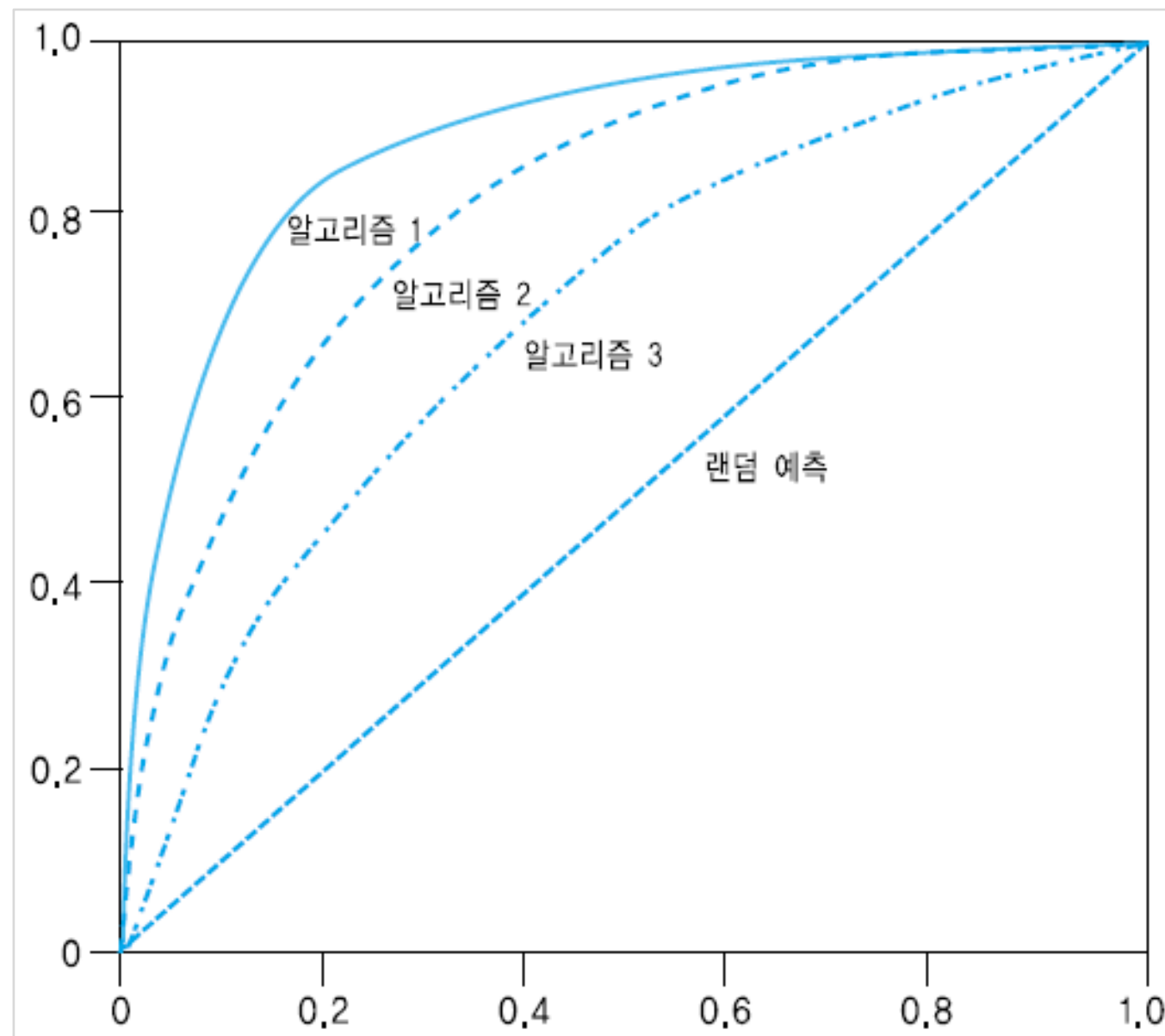


Bias-Variance



ROC, AUC

- ▶ 분류 모델의 ranking 성능을 평가
- ▶ ROC 그래프의 면적 AUC를 사용



분석도구가 중요한가?

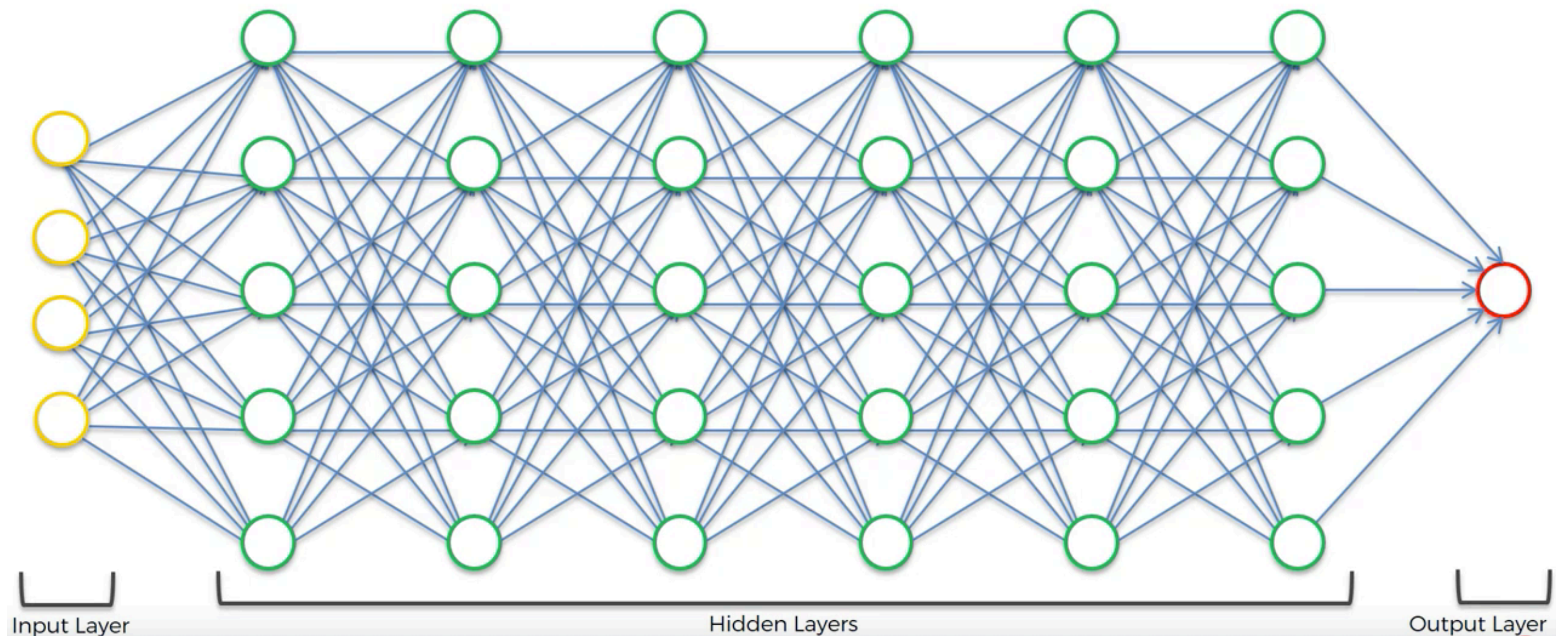
- ▶ 12회 SAS 마이닝 챔피언십
 - ▶ 문제: 특정 지역 병원의 폐업 위험성 예측 (심평원 후원)
 - ▶ 어떤 데이터가 중요한가?



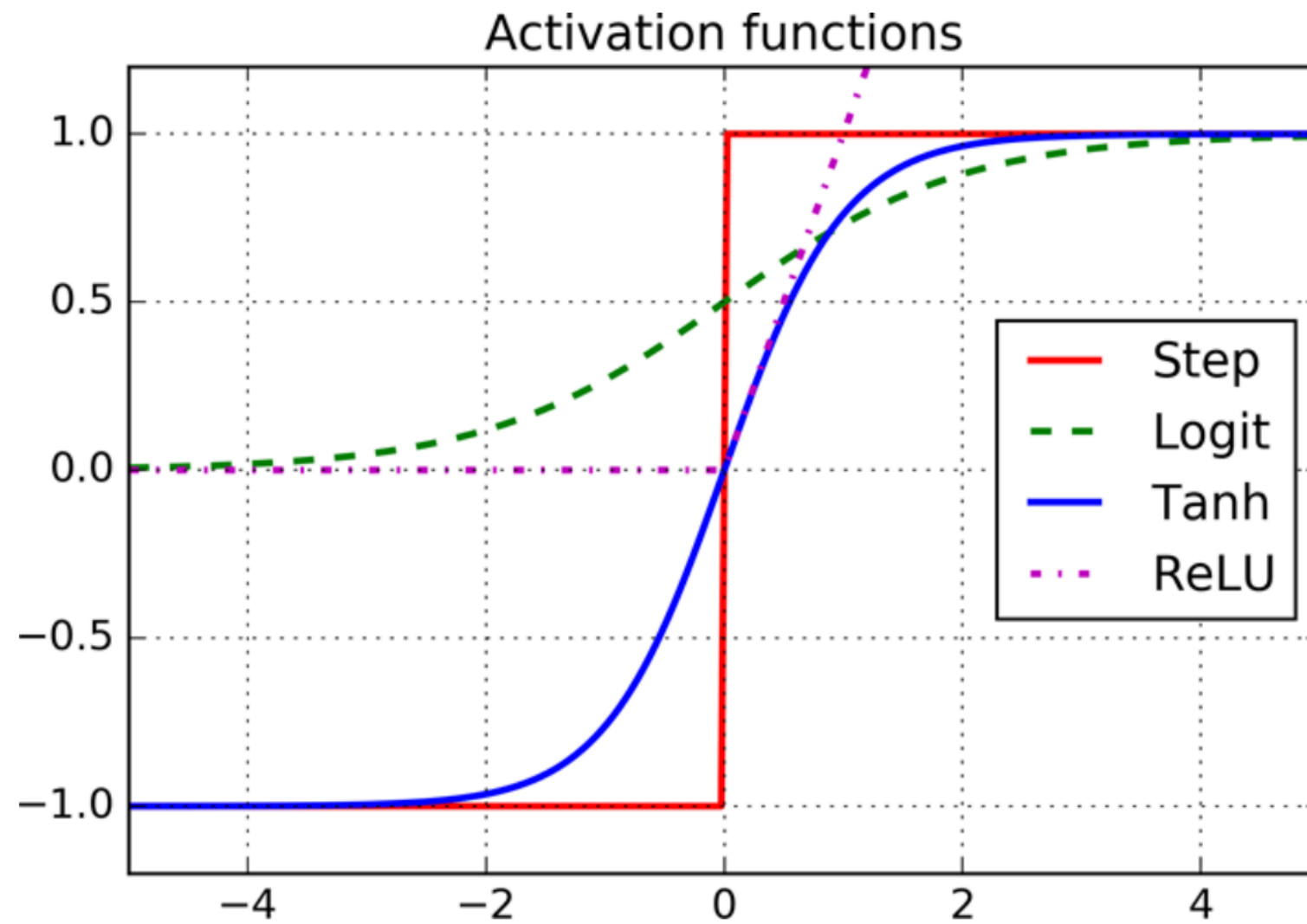
Deep Learning

Multi Layer Perceptron

- ▶ Fully Connected Network (전결합망)을 여러 층 누적
- ▶ Activation Function(활성화 함수) 도입
- ▶ 선형 모델보다는 개선되었으나 이미지 분석, 텍스트 분석의 성능은 기대에 미치지 못함

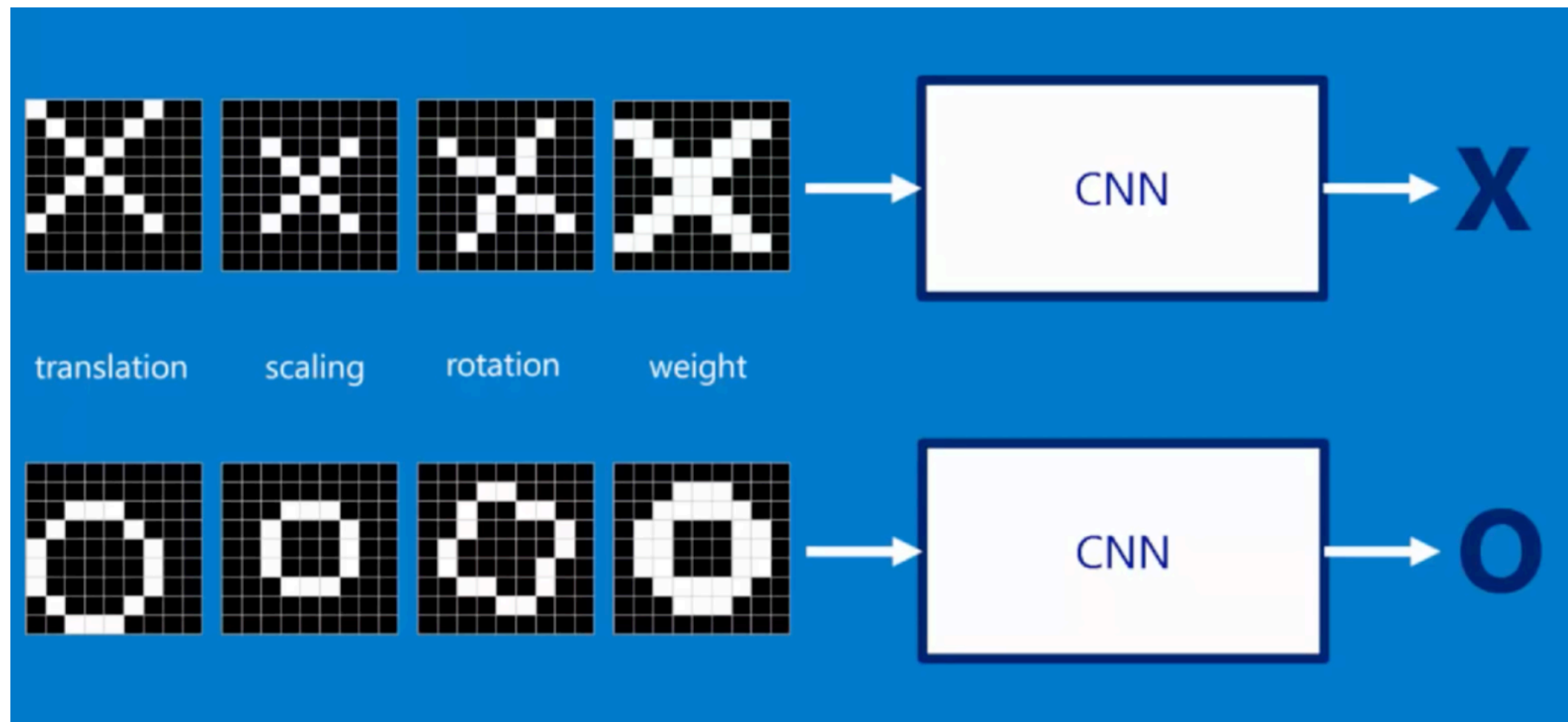


활성화 함수



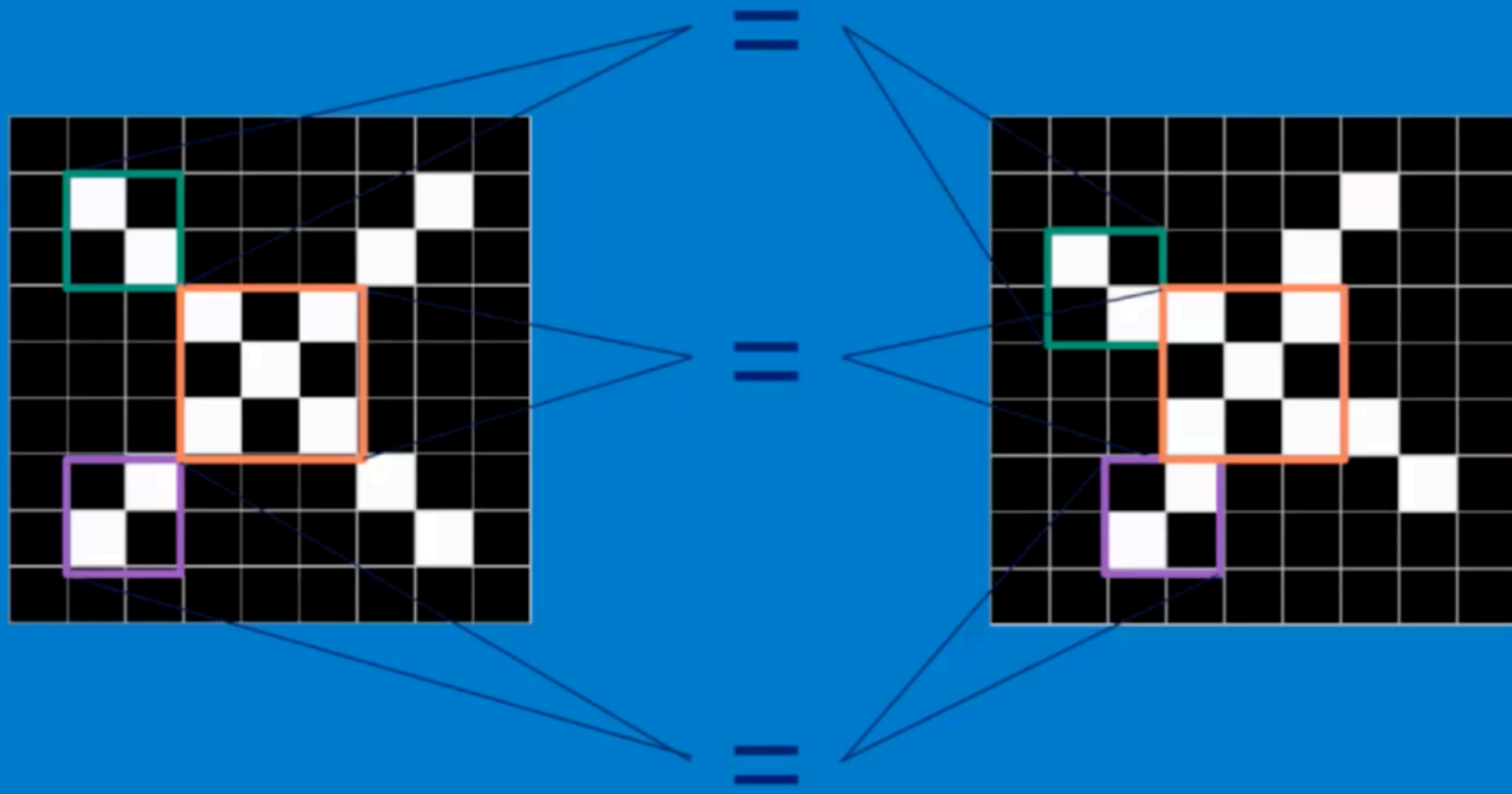
Convolution Neural Network

- ▶ 이미지 전체의 처리가 아니라 feature를 추출하고 이를 조합하여 이미지 판독의 성능을 향상

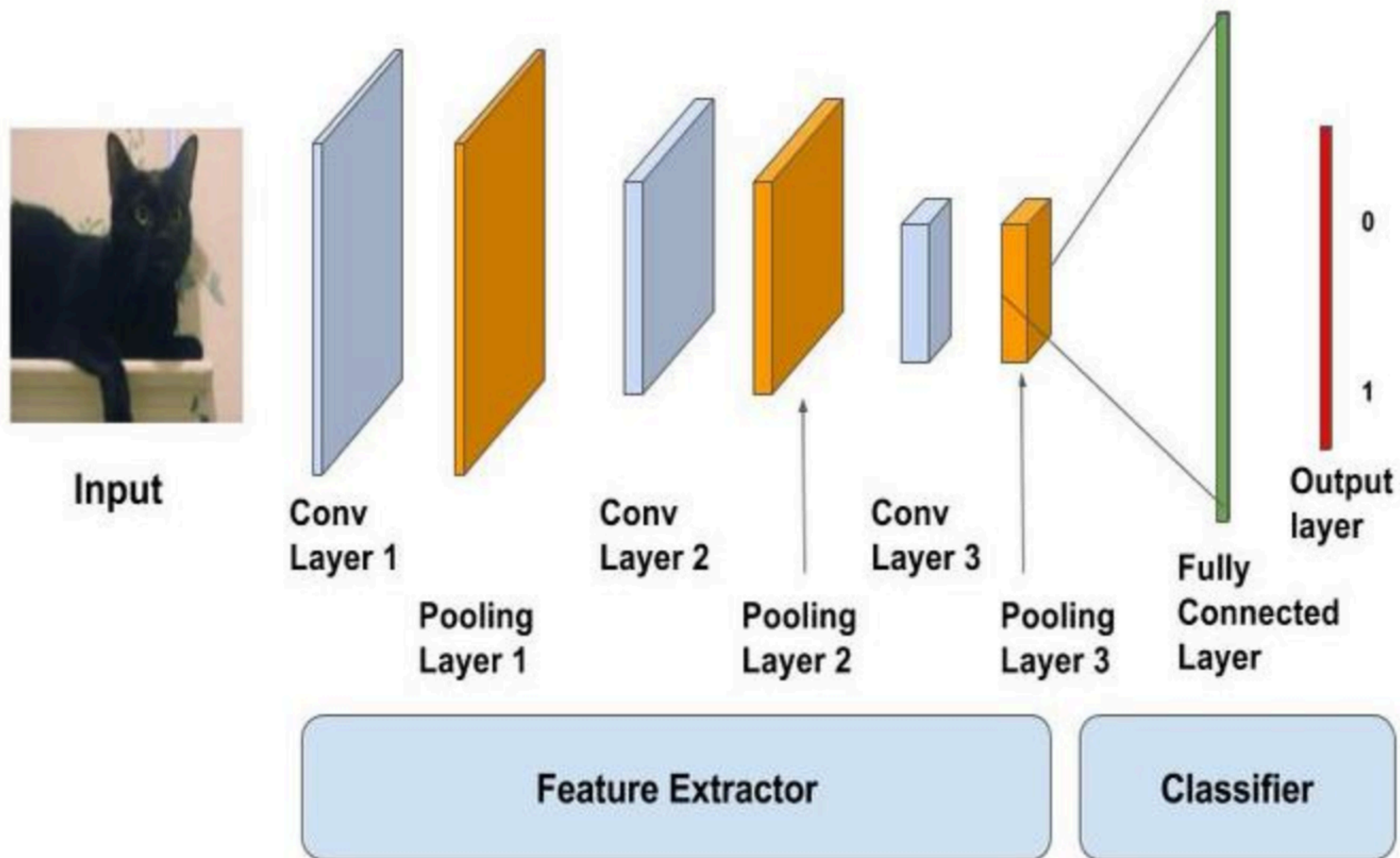


CNN

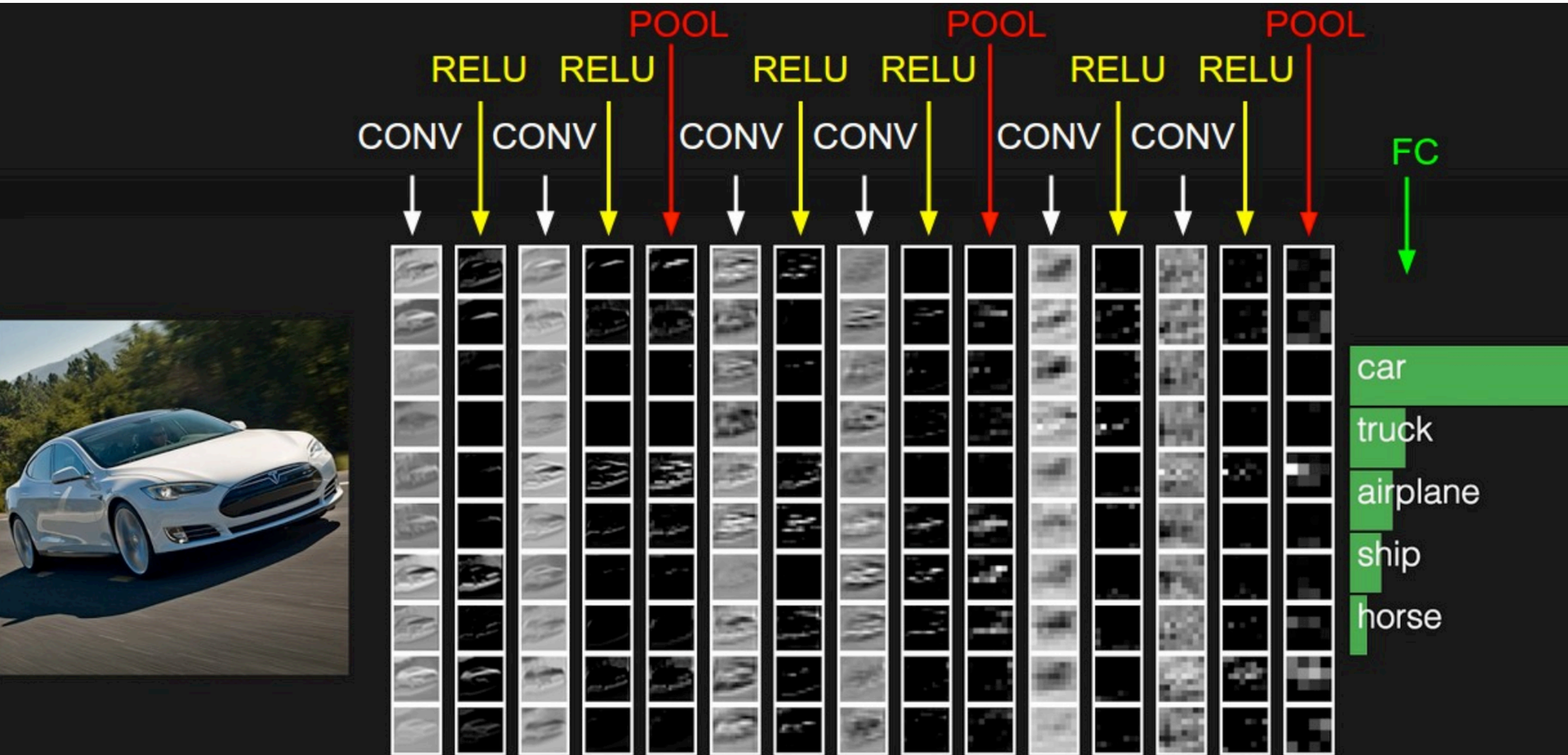
ConvNets match pieces of the image



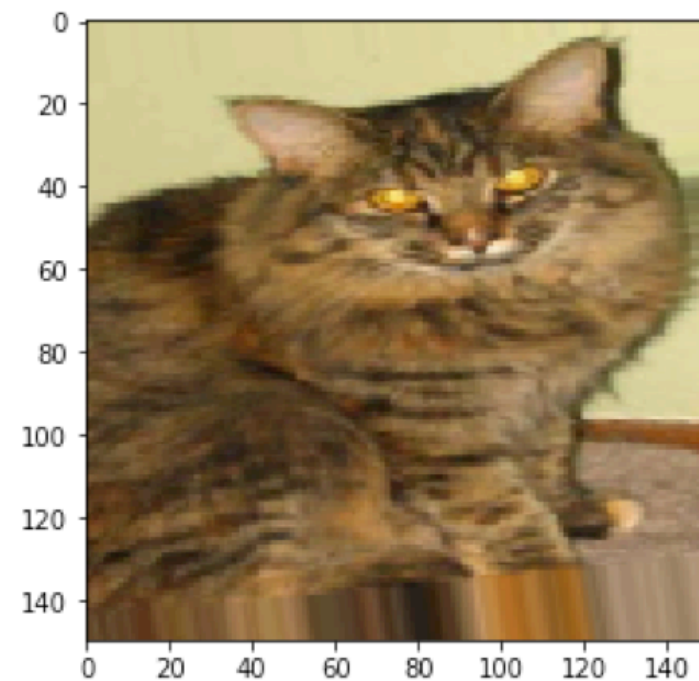
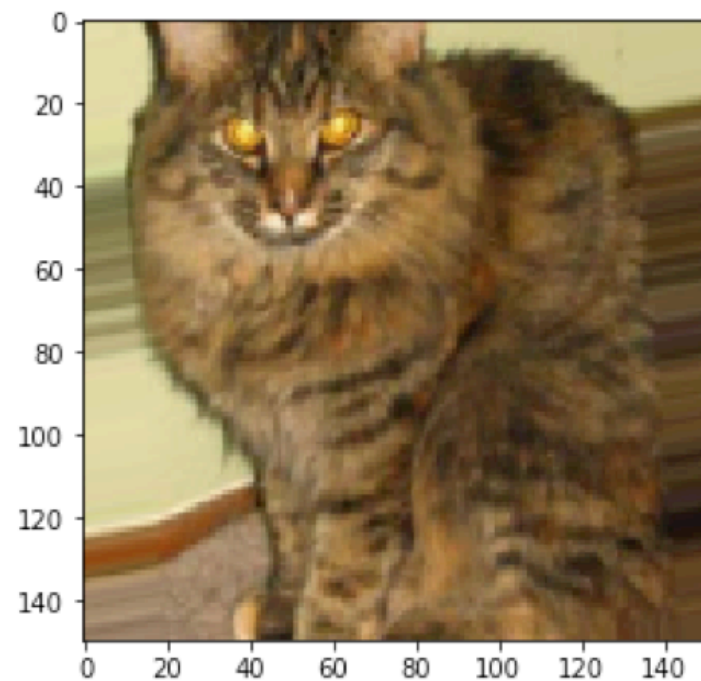
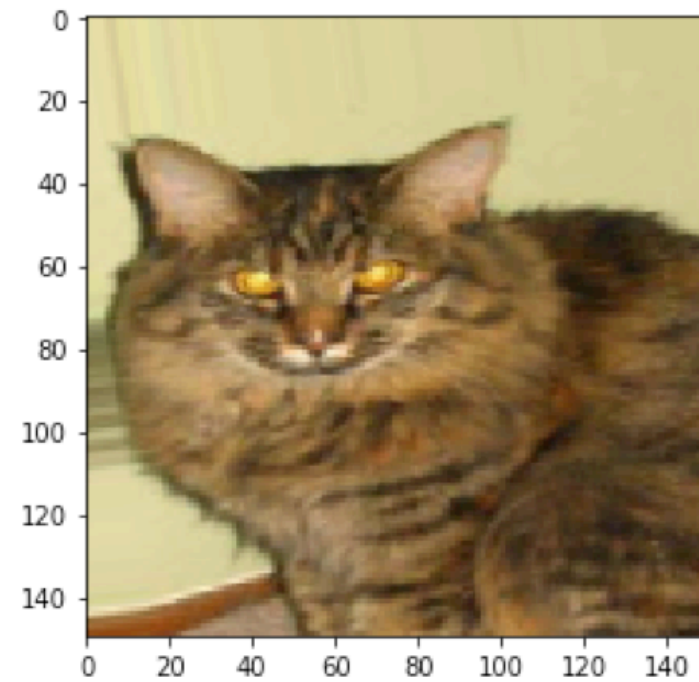
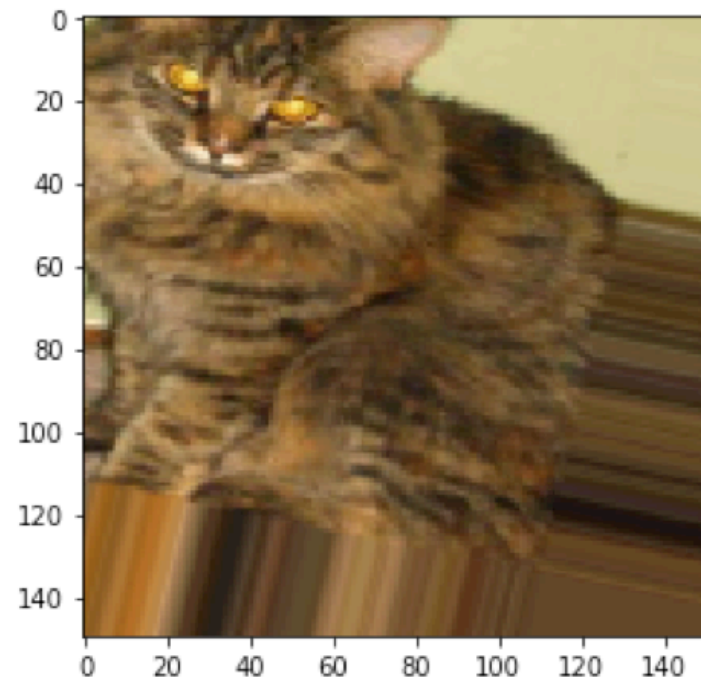
CNN 구성



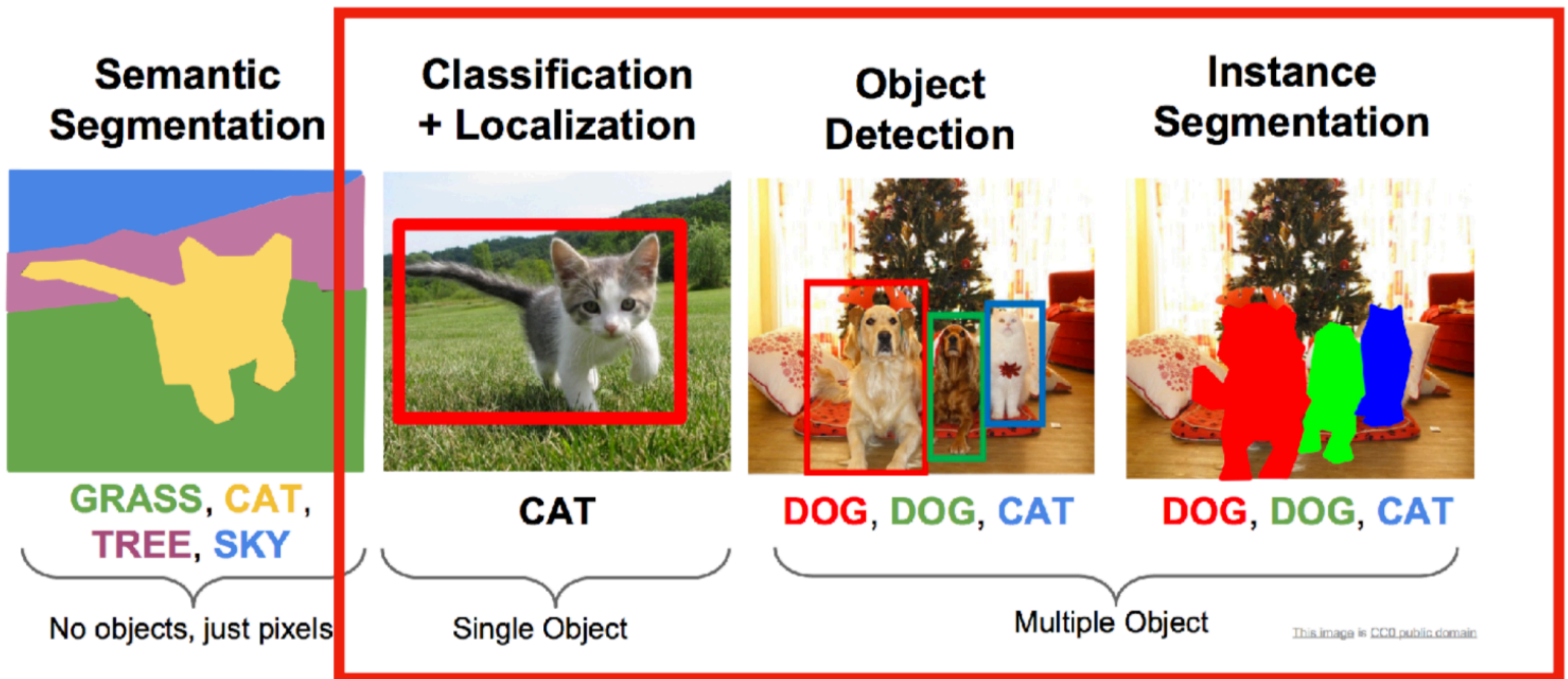
CNN 구성



Data Augmentation



Object Detection

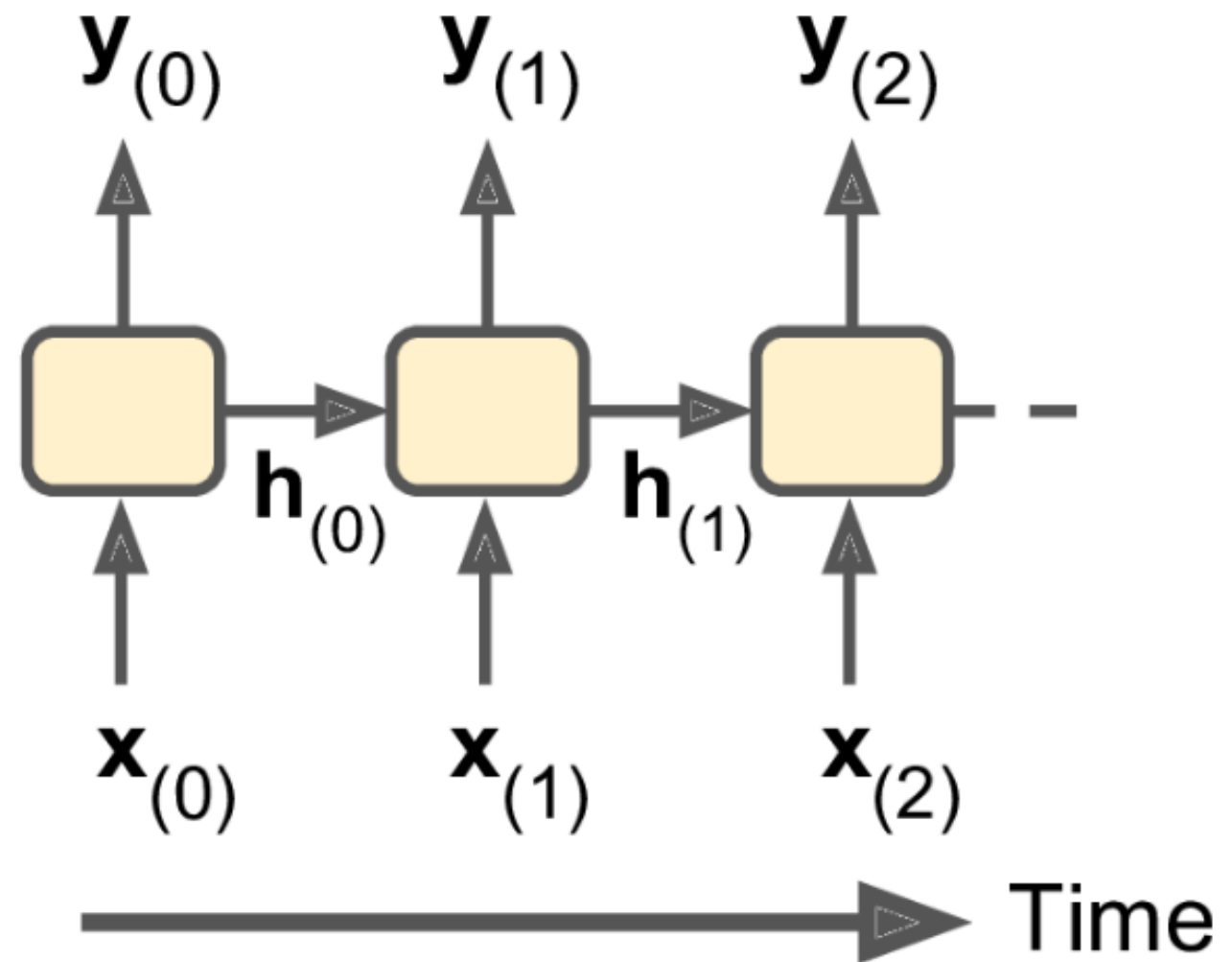
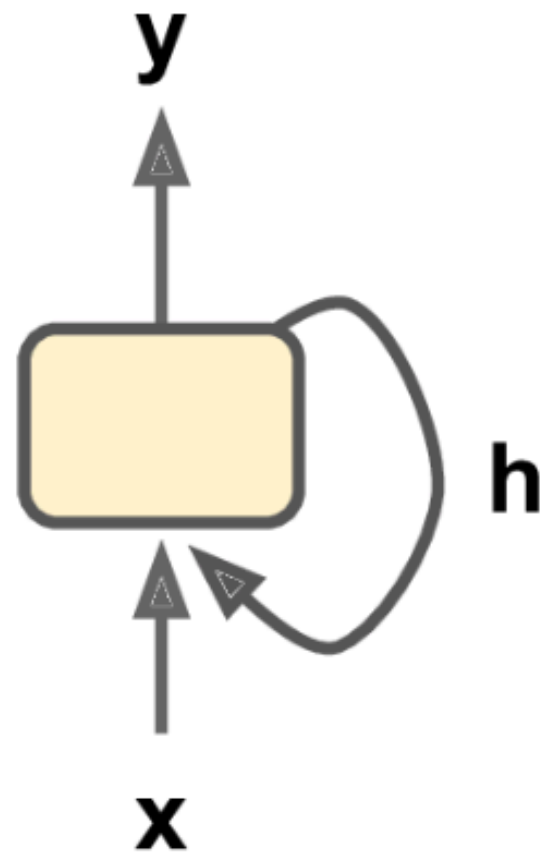


Mask R-CNN

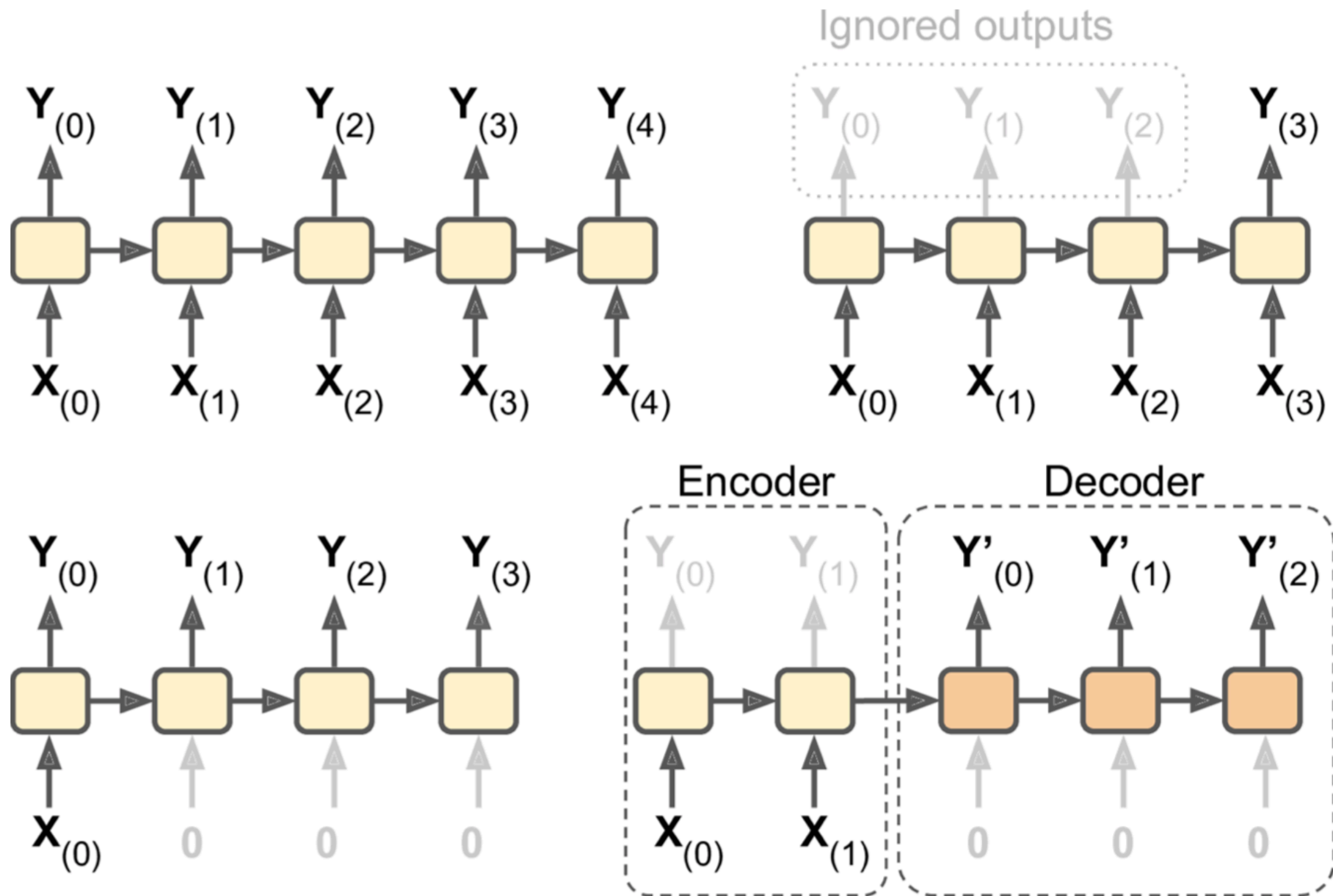


Recursive Neural Network

- ▶ 과거의 입력에 대한 상태 정보를 순환적으로 재사용

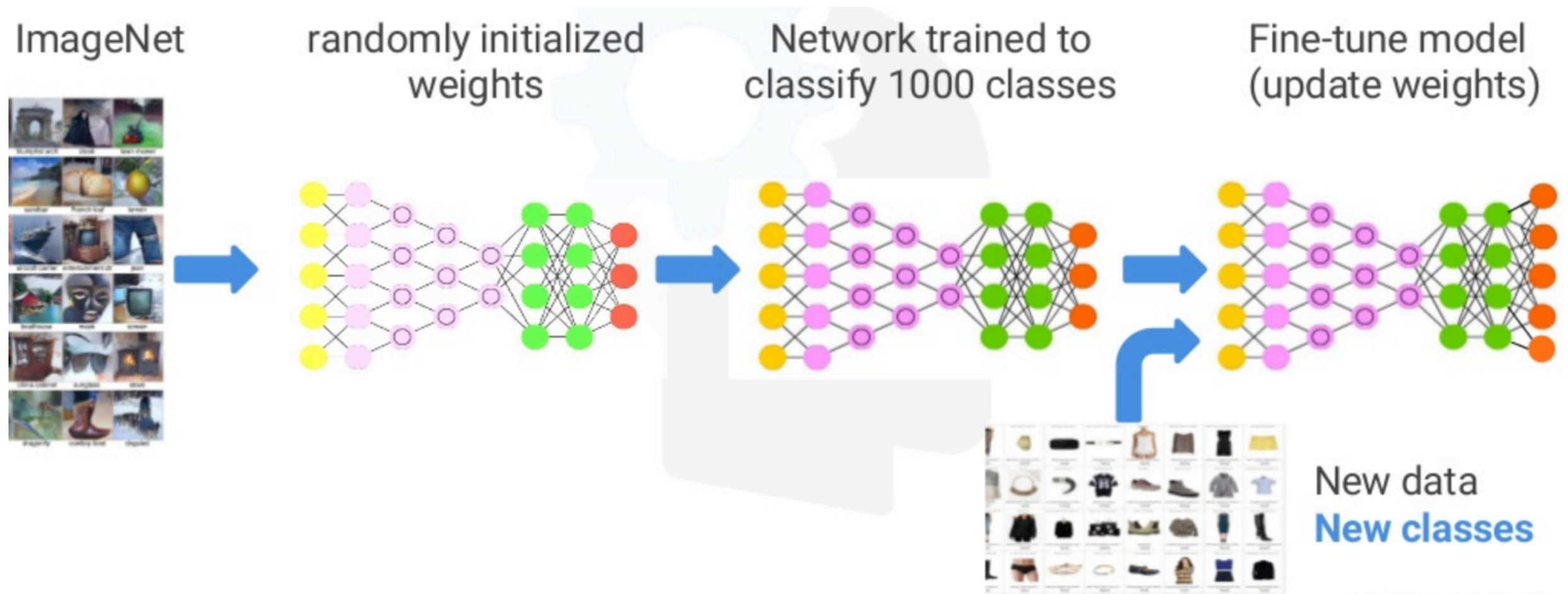


RNN 이용 유형



전이학습 (transfer learning)

- ▶ 다른 곳에서 만든 신경망 모델을 다른 분야에서 재사용



Block Chain

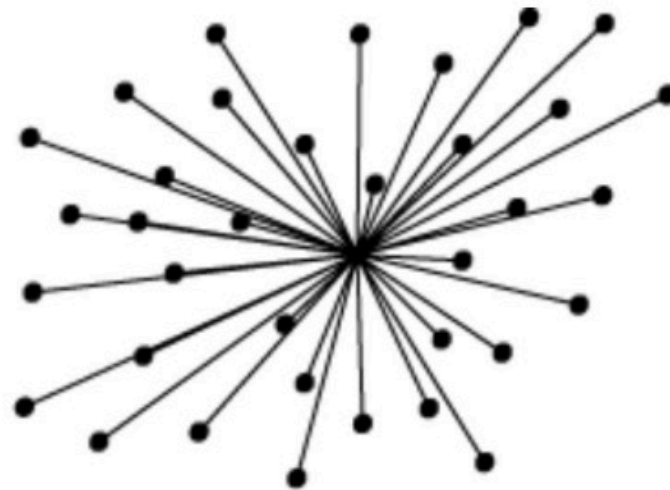
분산 시스템

- ▶ 중앙통제 시스템

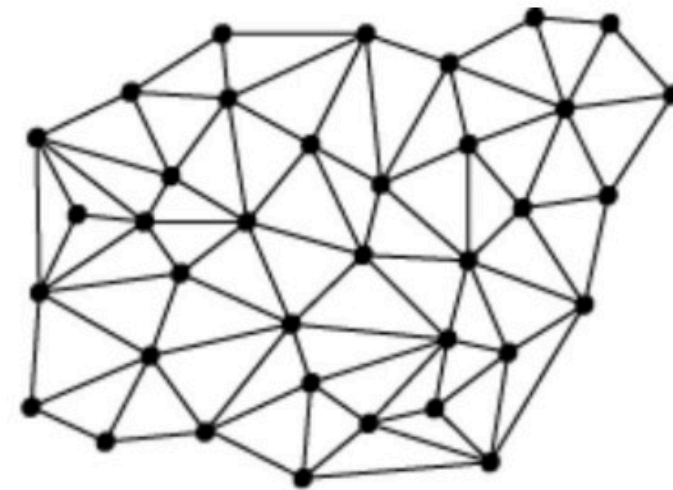
- ▶ 중앙 서버가 존재

- ▶ 분산시스템

- ▶ 연산을 분배할 수 있다 (비용절감)
 - ▶ 확장이 쉽다
 - ▶ 제어 오버헤드가 크다
 - ▶ 보안에 취약하다



centralised

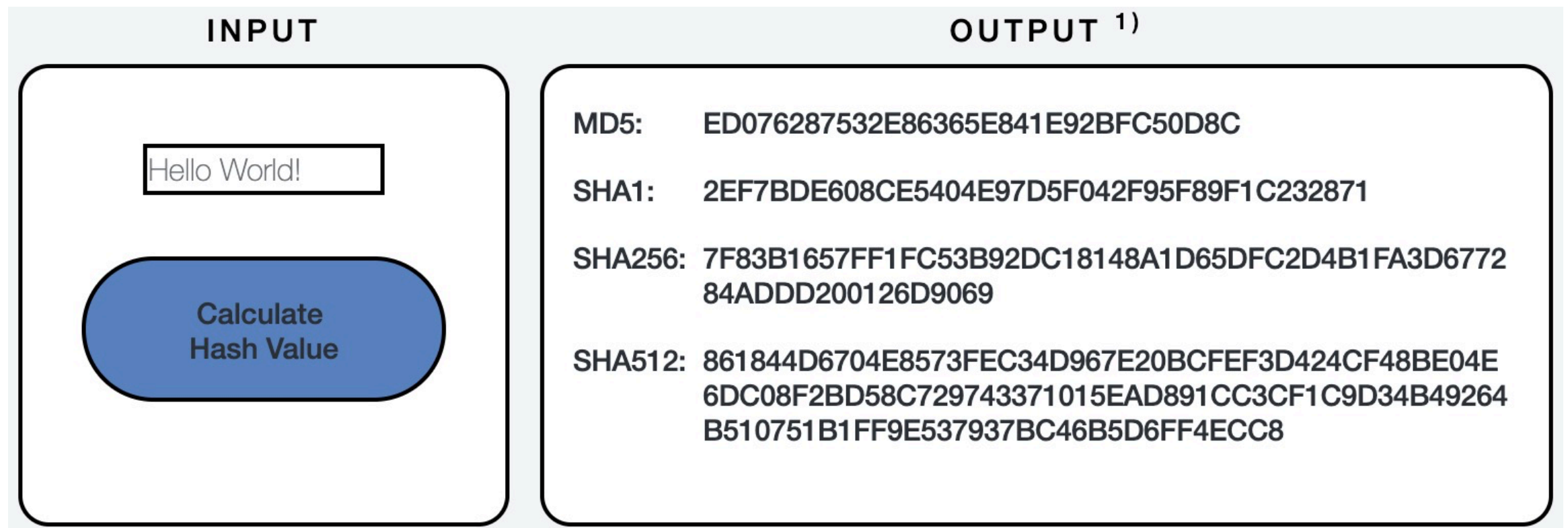


distributed

Hashing

▶ Hash function

- ▶ 임의의 입력에 대해 크기가 일정한 고유 패턴을 출력
- ▶ 동일 입력에 대해 동일한 해시값을 얻음
- ▶ 예측 불가능해야 함 (랜덤한 숫자)
- ▶ 동일한 해시값이 나올 확률이 매우 낮음
- ▶ <http://www.blockchain-basics.com/HashFunctions.html>



Chatbot

챗봇

- ▶ 대화형 UX

- ▶ 상담, 예약, 주문 등에 유용

- ▶ 성능은 아직 부족하지만 시간과 비용을 줄이는 효과는 있다.

- ▶ 유형

- ▶ 시나리오기반 대화
 - ▶ 딥러닝 기반 대화 - 수많은 대화 샘플로부터 학습

DialogFlow

▶ <https://dialogflow.com/>

Build natural and rich conversational experiences

Give users new ways to interact with your product by building engaging voice and text-based conversational interfaces, such as voice apps and chatbots, powered by AI. Connect with users on your website, mobile app, the Google Assistant, Amazon Alexa, Facebook Messenger, and other popular platforms and devices.

[Sign up for free](#)

Intro to Dialogflow

나중에 시청하기 공유

YouTube

Powered by Google's machine learning
Dialogflow incorporates Google's machine learning expertise and products such as Google Cloud Speech-to-Text.

Built on Google infrastructure
Dialogflow is a Google service that runs on Google Cloud Platform, letting you scale to hundreds of millions of users.

Optimized for the Google Assistant
Dialogflow is the most widely used tool to build Actions for more than 400M+ Google Assistant devices.

▶ 주욱 상단에 있는 "Go to console" 버튼을

AI in Medicine

의료 데이터 분석 분야

- ▶ EMR, 유전체, 임상 데이터 등을 분석
- ▶ 질병 예측
- ▶ 맞춤형 치료 추천
- ▶ 재입원률
- ▶ 입원중 사망률
- ▶ 퇴원시의 진단명

심혈관질환 예측

- ▶ 현재 가이드라인
 - ▶ 나이, 콜레스테롤 수치, 혈압, 흡연, 당뇨 등에 기반
- ▶ 영국 노팅햄대학교 연구 결과
 - ▶ 37만명의 EMR 분석
 - ▶ 인종차이, 정신질환, 경구용 스테로이드 복용 등이 주요 영향

유전체 데이터 분석

- ▶ 검사비가 200달러 수준으로 낮아짐
 - ▶ 혈액검사비 수준으로 낮아질 것
- ▶ 개인화된 의료, 정밀 의학의 시작
 - ▶ 개인별 적절한 치료 추천
 - ▶ 지금까지는 표준화된 의료 체계
 - ▶ 대규모 병원의 운영 비용 증가, 2차감염 문제 증가
- ▶ 맞춤형 제약
 - ▶ 현재의 “범용” 신약개발 비용은 계속 증가할 것
- ▶ 기업
 - ▶ Color Genomics
 - ▶ Illumina
 - ▶ 23andMe
 - ▶ Ancestry.com
 - ▶ CRISPR-Cas9

의료영상 분석

- ▶ 뇌, 폐, 복부, 심혈관, 유방, 뼈, 안구, 조직검사 병리 데이터 등
- ▶ MRI, CT, 초음파, x-ray, 유방촬영술, 안저사진, 피부사진 등
- ▶ 사례
 - ▶ Enlitic
 - ▶ Zebra Medical Vision

골연령 판독

▶ 뼈의 성숙도 측정

- ▶ 골연령을 예측하고 가장 유사한 표준 사진을 찾아준다
- ▶ 전문의가 활용할 경우 성능이 올라간다
- ▶ 판독 시간을 줄인다



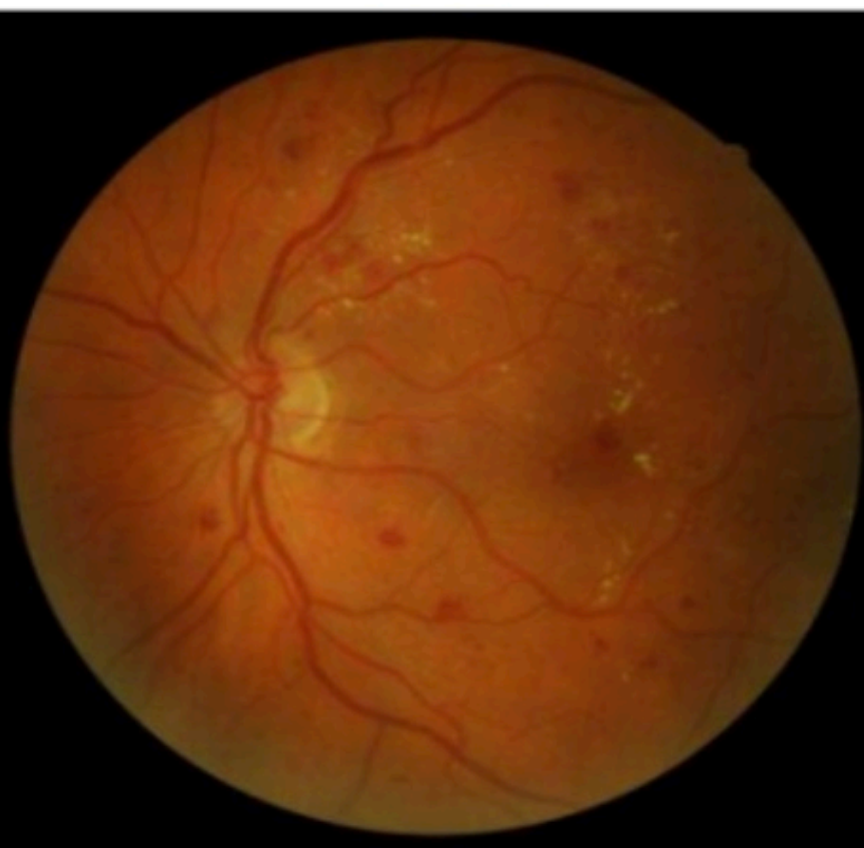
당뇨성 망막병증

- ▶ Diabetic retinopathy
- ▶ 사진을 눈으로 판독
 - ▶ 의사마다 판독 결과가 다른 경우가 많다

Healthy Retina



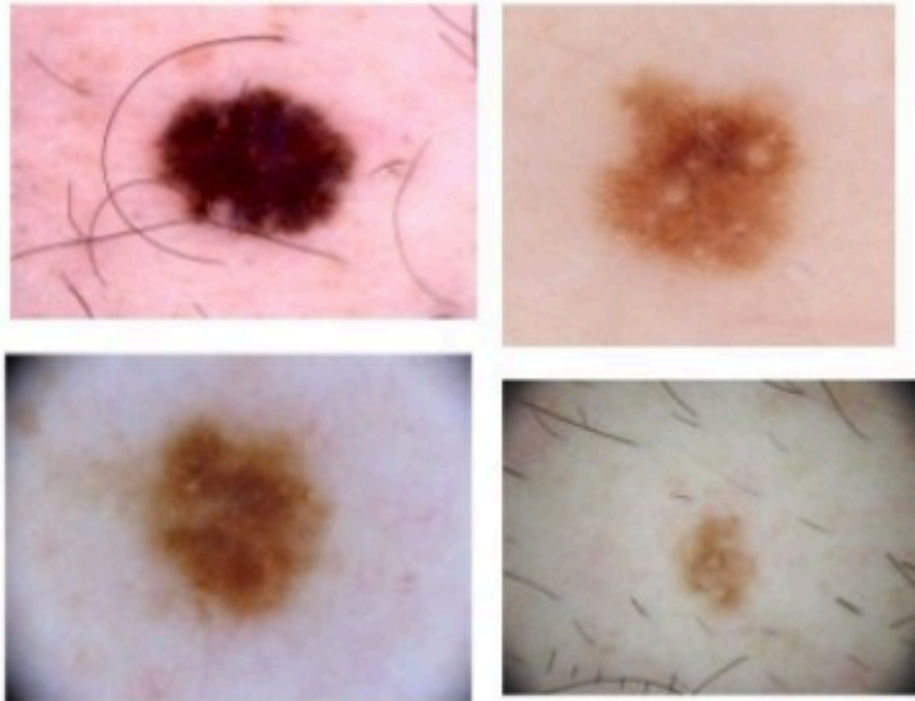
Diabetic Retinopathy



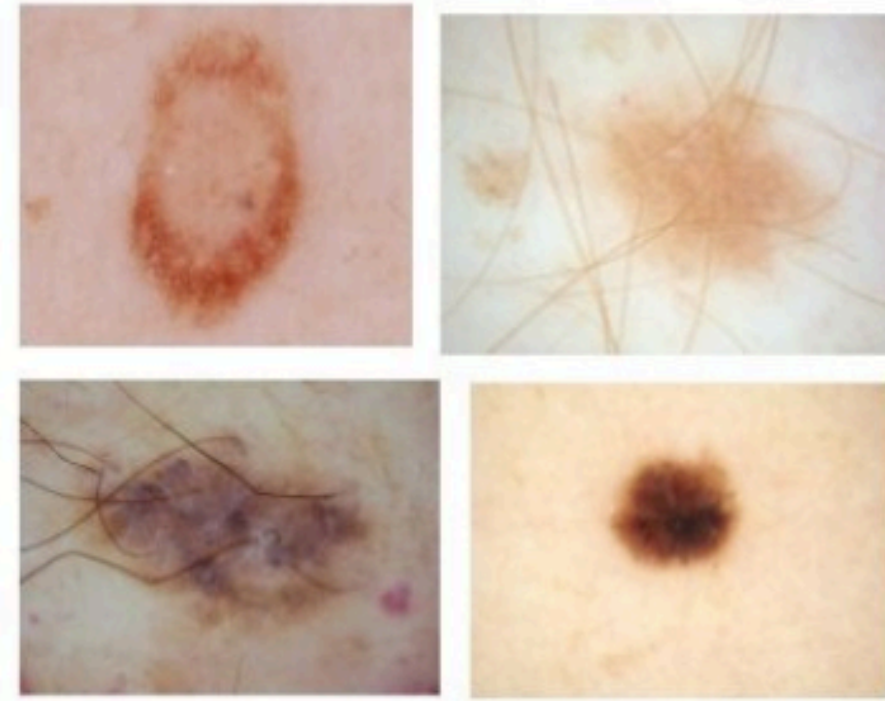
피부암

- ▶ 스탠퍼드대학에서 피부암 진단 모델 구현
- ▶ 2017. 12 네이처에 발표
- ▶ 조기 발견이 중요
 - ▶ 흑색종(melanoma)의 경우 다른 점, 검버섯, 사마귀 등과 구분이 어려움
 - ▶ 딥러닝 모델로 AUC 값이 0.9이상을 얻음
- ▶ 스마트폰 앱
 - ▶ Dermatoscope 부착도 가능

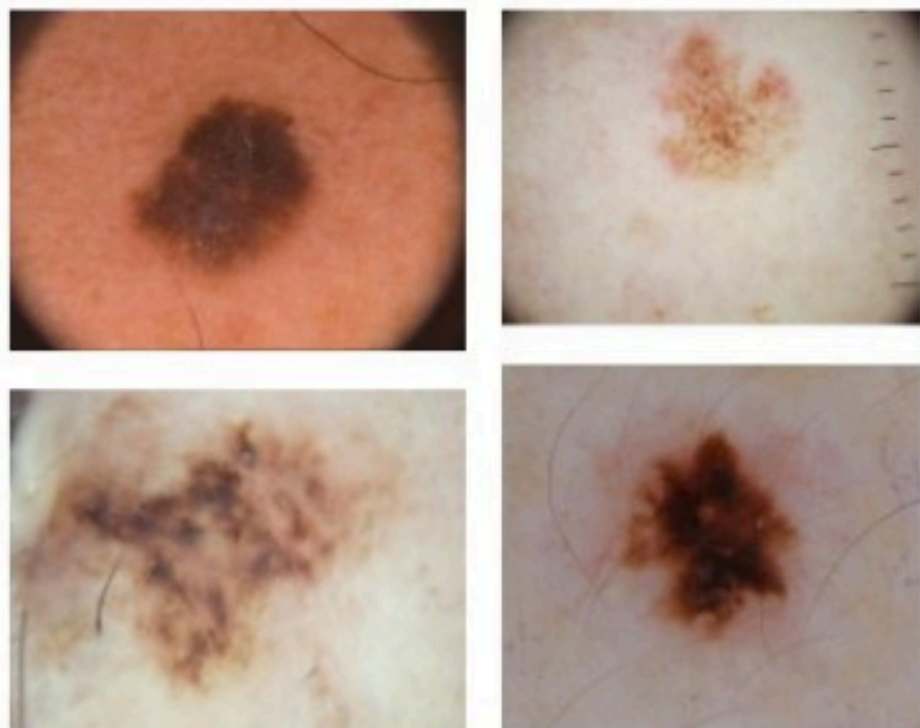
True Positive



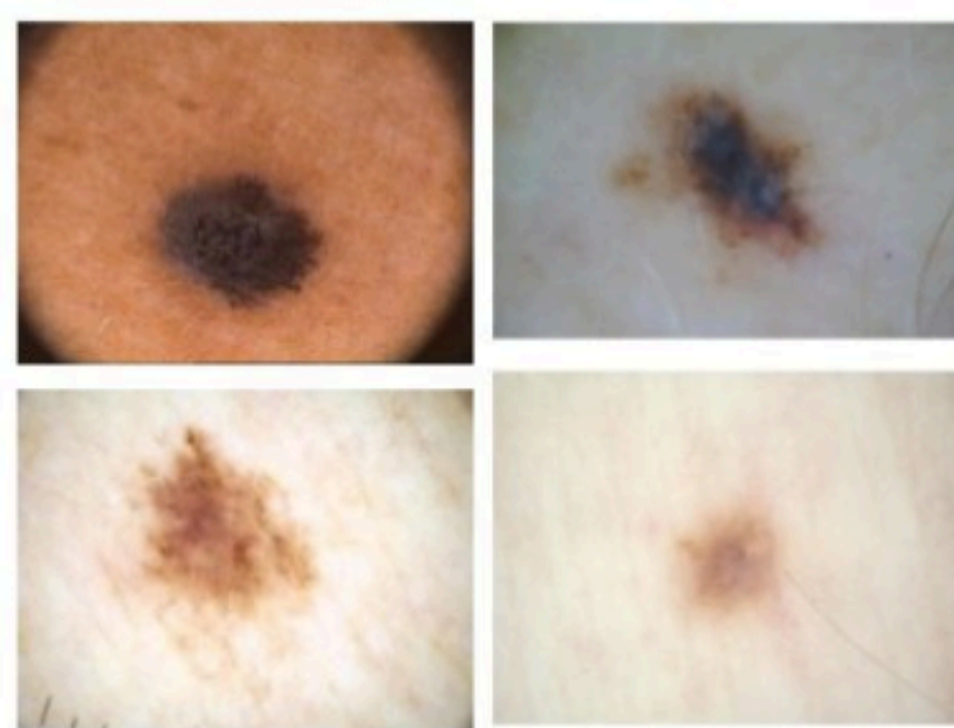
False Negative



False Positive

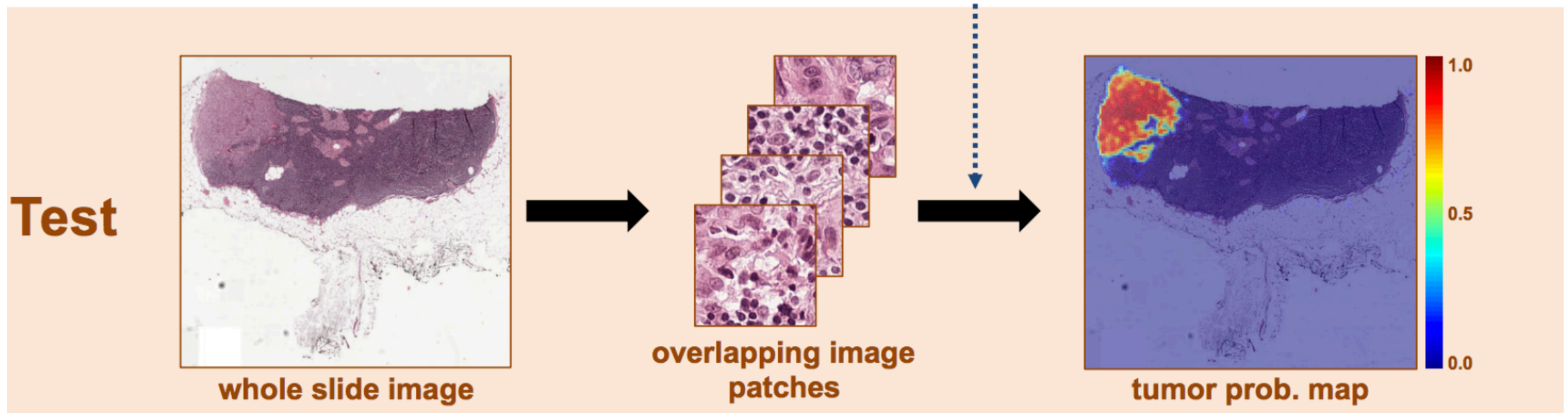


True Negative



병리학

- ▶ 하버드대 Andrew Beck 교수팀
 - ▶ 유방암 겨드랑이 림프절 전이 여부와 위치 예측



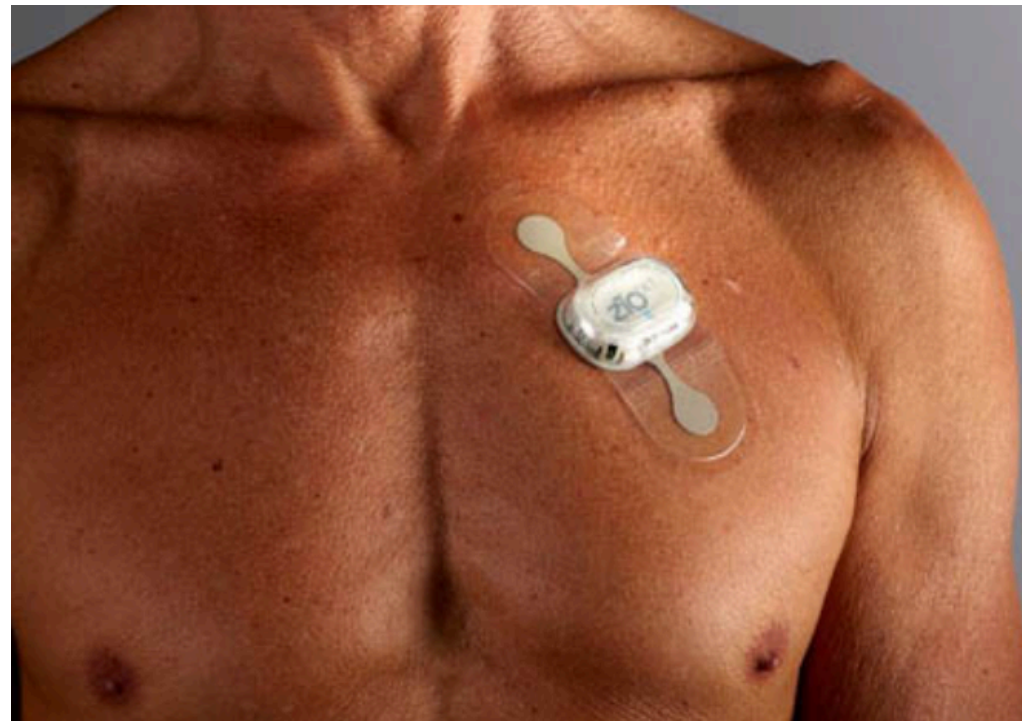
실시간 모니터링

- ▶ 신생아 중환자실에서 패혈증 예측
 - ▶ IBM, 온타리오대- Project Artemis
 - ▶ 체온변화, 자극에 대한 반응 감소, 혈압저하, 호흡수 증가 등 분석
- ▶ 존스 홉킨스대
 - ▶ 일반 중환자실의 패혈증 쇼크 예측 및 경고



부정맥 예측

- ▶ 서울아산병원 심실빈맥을 1시간 전에 예측하는 모델 개발
- ▶ Cardiogram은 애플워치의 심박센서만으로 심방세동과 심방조동을 측정
- ▶ 스탠퍼드대에서 부정맥 예측 모델 개발 (2017.7)
 - ▶ 34층의 딥러닝 모델 사용
 - ▶ 단일 채널의 심전도만 측정 지오판치 (2주일간 동작)



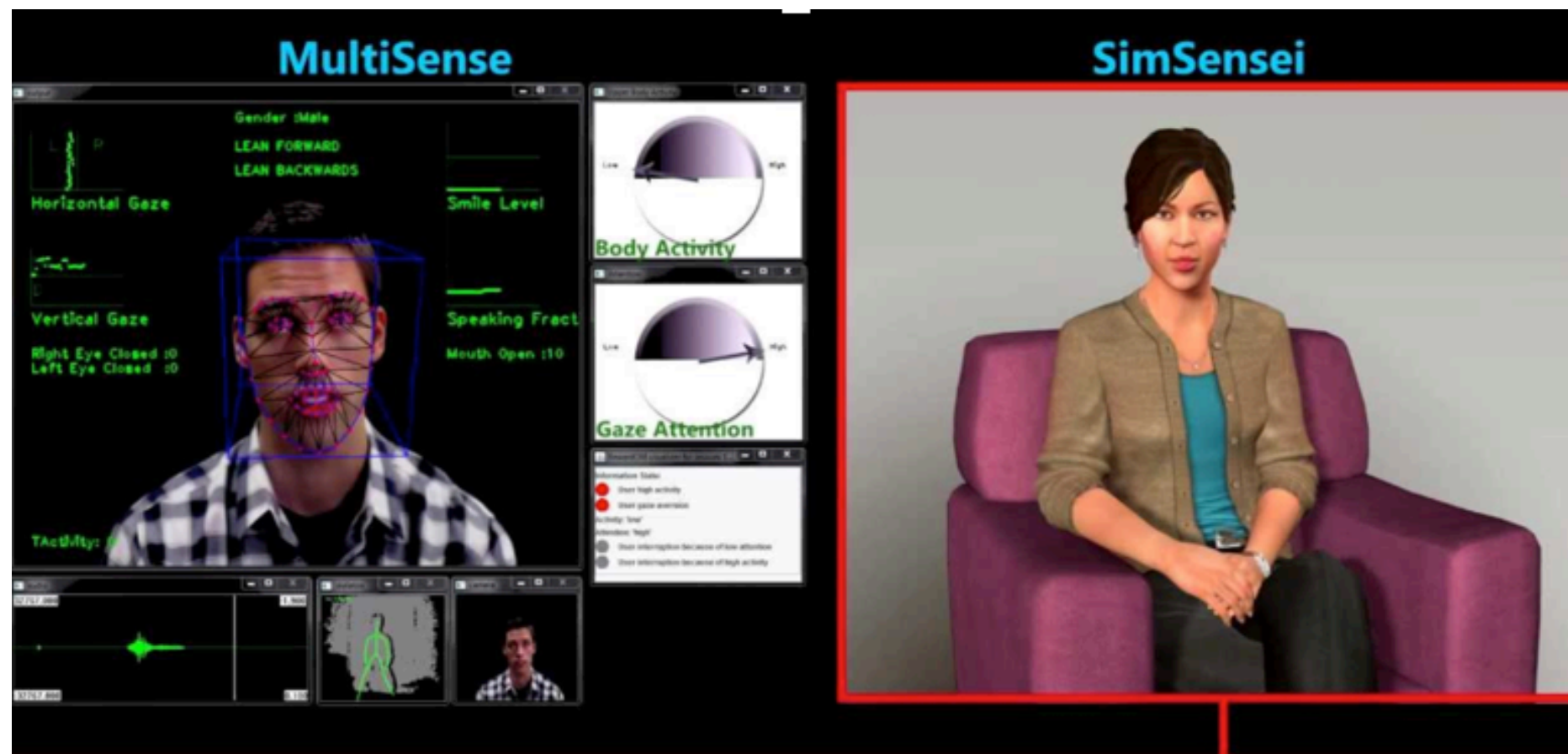
Sedasy

- ▶ 존슨앤존슨의 자동 수면 마취기 (2013)
 - ▶ 마취약 자동 주사
 - ▶ 심박수, 산소포화도, 심전도, 혈압을 보고 투약량 조절
 - ▶ 수면 내시경 의료비를 1/10로 낮춤
 - ▶ 더 빠르게 회복,
 - ▶ 저산소증문제도 적게 겪음
- ▶ 2016년 시장에서 철수



SimSensei

- ▶ 인간 의사보다 AI의사에게 더 솔직하게 상담
- ▶ 서던 캘리포니아대에서 개발
 - ▶ 가상의 여성 상담사
 - ▶ 환자의 시선, 미소, 머리 움직임, 표정 분석



정신과

- ▶ 목소리로 감성 분석 (Beyond Verbal)
 - ▶ 앱 Moodies
- ▶ 언어적 요소 linguistics 뿐 아니라 비언어적 요소 acoudtics도 분석
- ▶ 조현병, 자살위험군 예측
- ▶ 우울증: 6초간 목소리 분석으로 AUC 0.93 (PureTech Health)
- ▶ 조울증 90% 진단

Robot

► Da Vinci



방향 - 예방의학

- ▶ Predictive Medicine – 치료 받으러 오는 사람을 줄이고 평소의 건강 관리자를 늘려야
- ▶ Precision medicine (정밀의료) – 유전체, 임상정보, 생활 환경 및 습관(Lifelog)을 분석하여 맞춤형 예측 의료
- ▶ Unity Biotechnology – 노화조직을 젊게 만드는 약물 개발
- ▶ Elysium Health – 20대부터 감소하는 조효소 NAD+를 늘리는 방법 연구

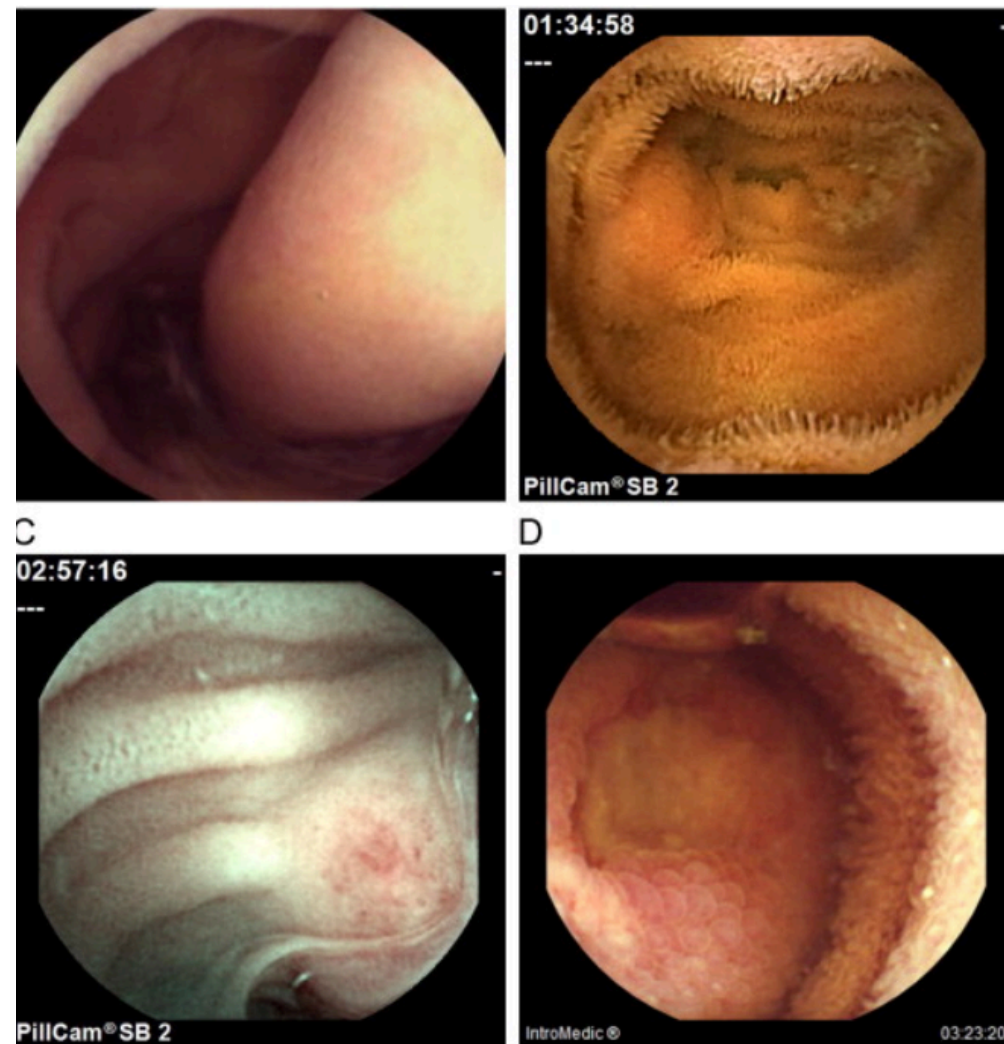
방향 - AI와의 협업

- ▶ 사람은 특이도가 매우 높다 (암이 아닌데 암이라고 판단하는 확률이 매우 낮다)
- ▶ AI에게 특이도의 기준을 완화하면 민감도를 매우 높일 수 있다 (암을 절대 놓치지 않는다)
- ▶ AI가 고른 후보를 사람이 확인하면 서로의 장점을 살릴 수 있다
- ▶ 정확도 뿐 아니라 업무 부담도 줄인다

IOT

캡슐 내시경

- ▶ 촬영된 사진이 너무 많아 판독에 시간이 오래 걸린다
- ▶ 1초에 2~3장 송신

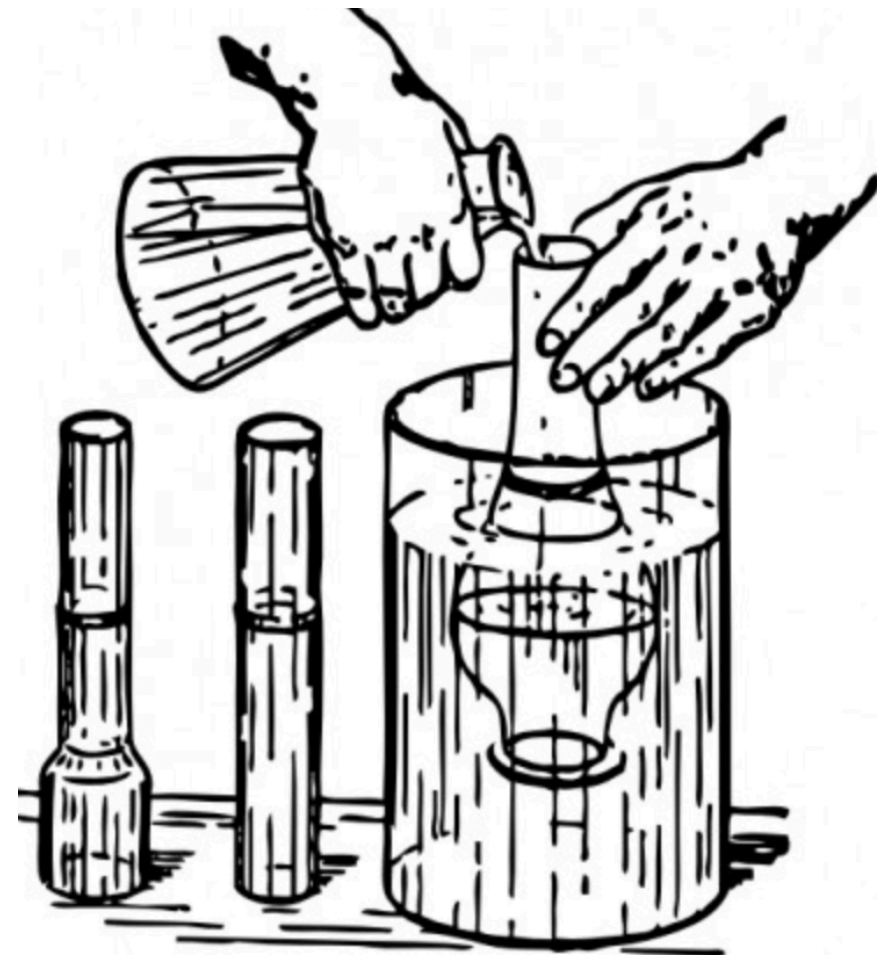


IOT 활용 사례

- ▶ Warby Parker – 휴대전화로 시력 검사
- ▶ Spruce – 피부과 의사 원격 상담과 처방
- ▶ UroSense – 소변 분석
- ▶ OrthoSensor – 인공 무릎관절 모니터링
- ▶ HealthPatch(Vital Connect) – 건강정보 송신
- ▶ OvulaRing(VivoSenseMedical) - 여성의 가임기를 알려줌

Data, Data, Data

- ▶ 가치 있는 데이터를 확보하는 전략이 핵심



Augmented Intelligence

- ▶ **AI가 잘 하는 영역**
 - ▶ 기억 memorization
 - ▶ 처리 transactions
 - ▶ 예측 prediction
 - ▶ 추천 recommendation
- ▶ **사람이 잘 하는 영역**
 - ▶ 감성 능력 emotive capability
 - ▶ 창의성 creativity
 - ▶ 종합 능력 (철학, 윤리) generalization
- ▶ **협력 지능(collaborative intelligence)**
 - ▶ AI를 잘 활용하는 능력

What to do

AI 도입전략

▶ 사람

- ▶ 내부 전문가 및 팀을 양성해야
- ▶ 여러 영역의 협업이 필요하다
- ▶ 전문 파트너 사를 신중하게 선택해야 한다

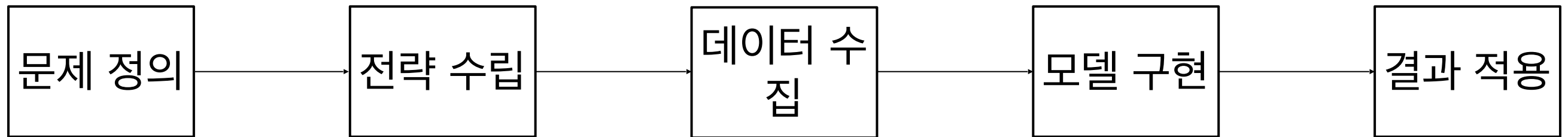
▶ 데이터

- ▶ 대부분의 시간과 에너지가 데이터 준비에 사용 60~80%
- ▶ 다른 곳에 없는 고유한 데이터가 가치 있는 자산
- ▶ 외부 데이터, 공개 데이터, 고객데이터의 활용 전략

▶ 파일럿 프로젝트

- ▶ 6~12 개월 내에 성과가 나는 과제 도출
- ▶ 경쟁력 있는, 차별화된 사업을 추진하고, 일반 기술은 도입

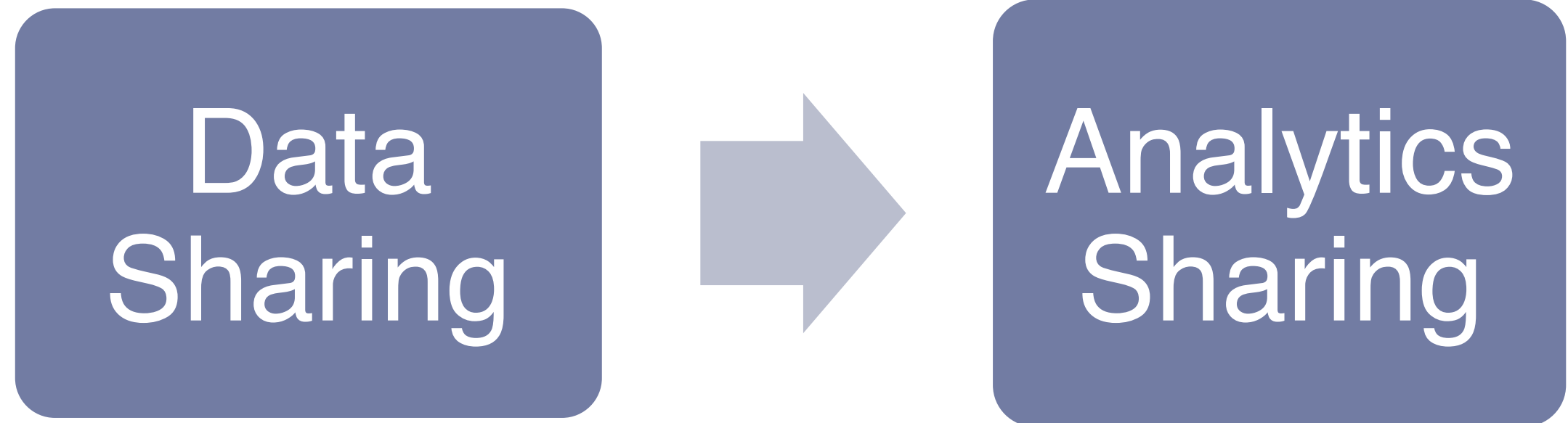
Project Process



- ▶ 문제 정의 - 해결하려는 문제를 명확히 정의하는 것
- ▶ 전략 수립 - 어떤 데이터를 누가 어떻게 사용할지를 정함
- ▶ 데이터 수집 - 머신러닝에 필요한 데이터를 수집하는 것
- ▶ 모델 구현 - 머신러닝 모델을 구현
- ▶ 테스트(결과 적용) - 모델을 실제 상황에 적용하고 성능을 개선
- ▶ 리포팅 - 관계자의 이해와 설득

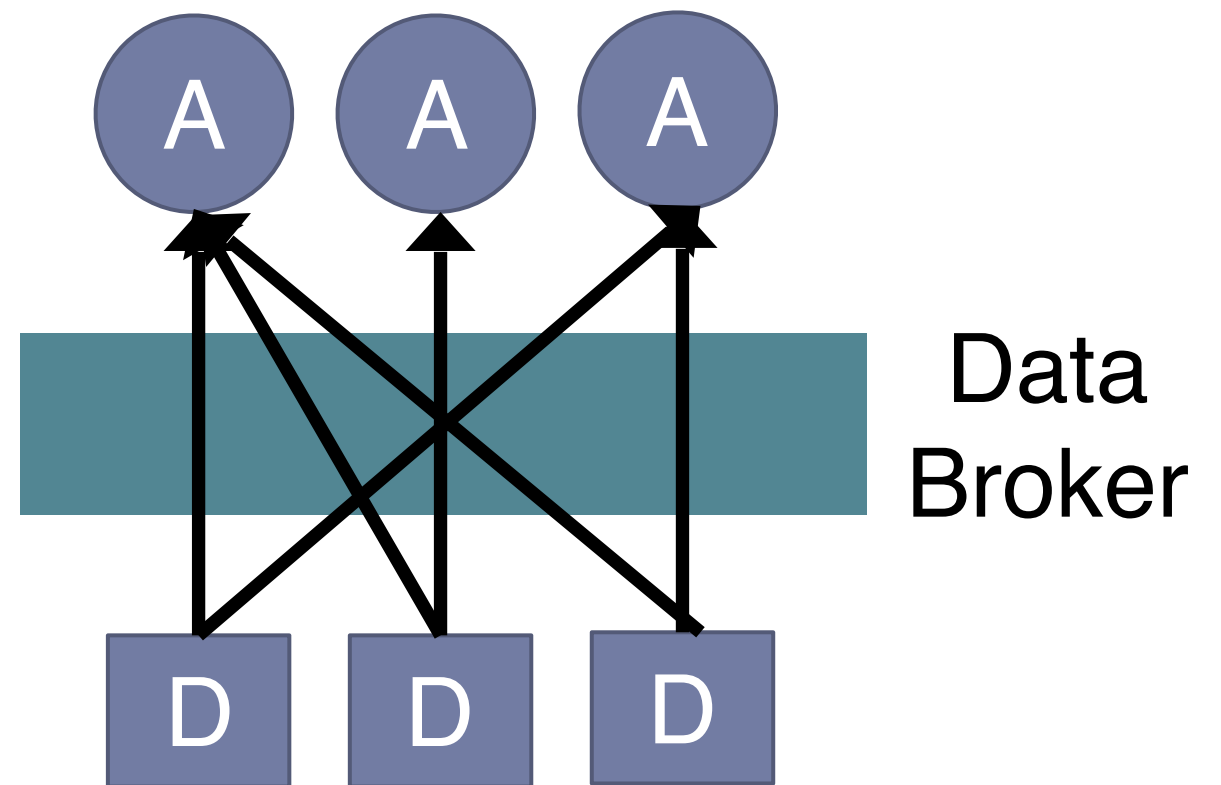
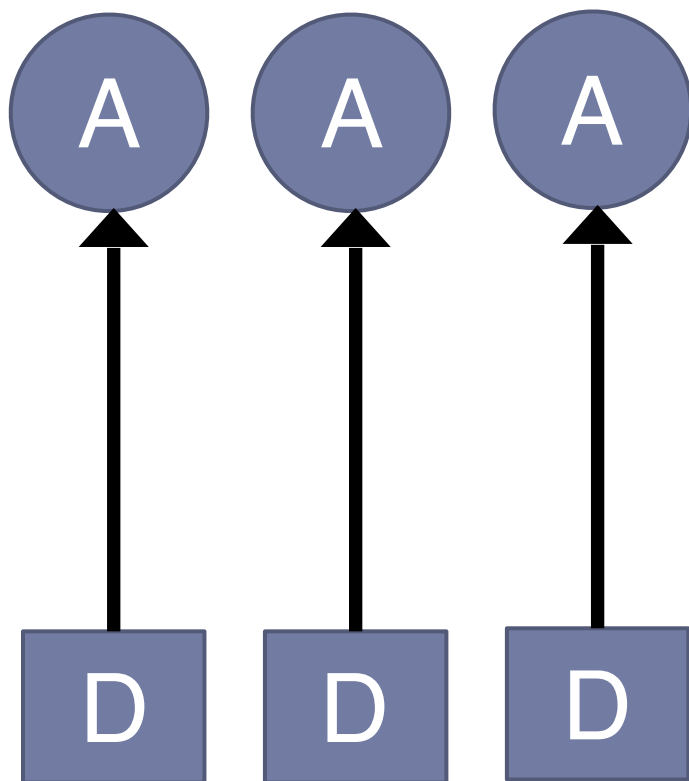
Suggestion

- ▶ 데이터를 확보하는 궁극적인 목적은 분석임



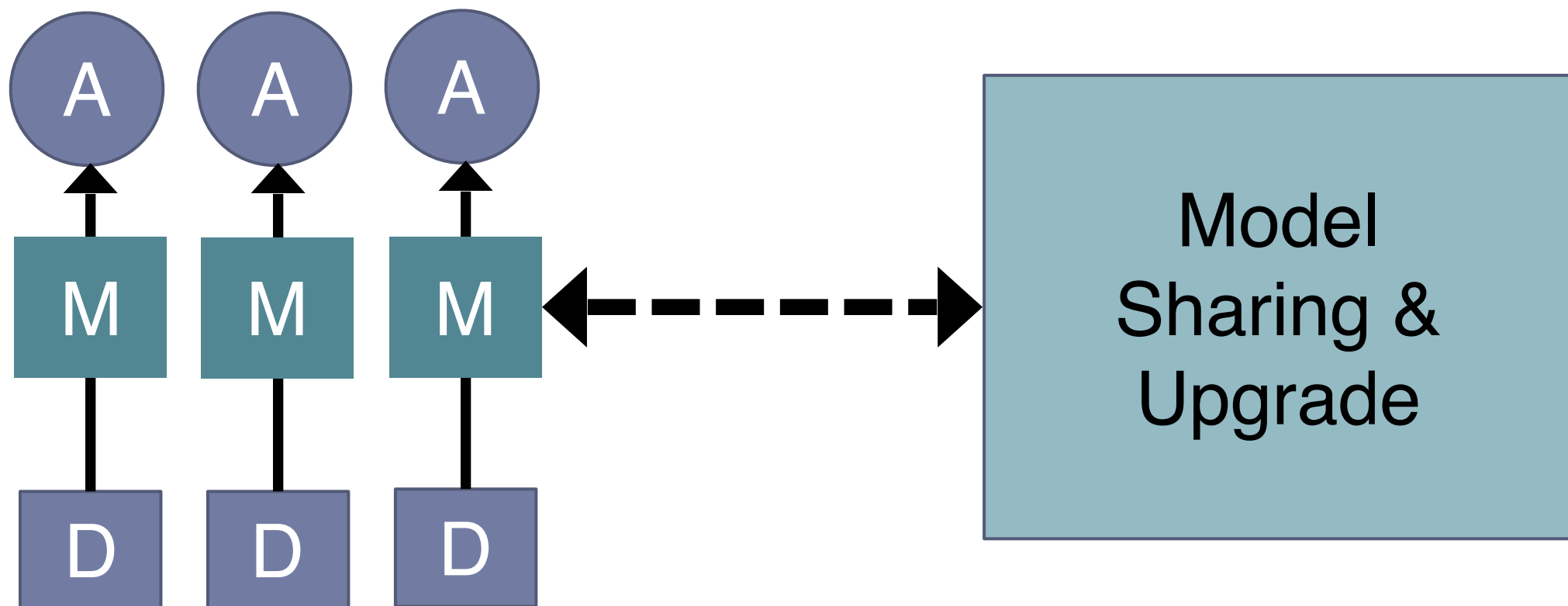
Data Sharing Approach

Traditional Analysis



Analytics Sharing

Transfer Learning



Conclusions

- ▶ AI will be used everywhere SOON
- ▶ Key element of AI is **machine learning**
- ▶ Gather and use **Valuable** Data
- ▶ Lead a disruptive project – “**Analytics Sharing**”

Q & A