

***BigData\_MachineLearning\_DeepLearning***

# 머신 러닝

# Machine Learning

**컴퓨터에게  
뭔가를 가르쳐서  
잘하게 한다.**

## 맞춤 동업삼



☞공공선광고를듣는대로해보  
나.....☹



(Top7랭킹) 몰입도,충족성,퍼  
미드 추천 7선



송지효 시스루 미상한 눈으로 쳐  
보는 이광수

연봉 PD  
프로수 57.4만회, 2주 전



최익 TV  
주최수 2.4천회 · 3시간 전



노는 애틋한 커서 뭐가 될까? 밀진  
직업 TOP 5

이슈모드급  
조회수 62만회 · 8개월 전



블래시 로얄 애니메이션 해골 병사  
의 최후

호모노  
조희수 9.6만회, 1주 전



사냥하러 가볼까! (The Best Hunter DECK) [클래시로드]  
JUNE  
조희수 2,7만원 - 7시간 전



동물실험 수정판.avi  
SukJungOh  
조회수 160만회 • 6년 전



돌아서면 누가 가장 예뻐서 것 같습  
니제?  
DORI TORI  
조희수 51.3만회 · 1개월 전



인간의 욕심 때문에 천화가 아닌  
목화를 길정할 동물

영킹도어  
조희수 35.9만회 + 2개를 전



마인크래프트] 주인은 에메랄드 진  
을 좋아할까? - 꿀잼 주연 실험 (과  
게임하는 올빼미 TV  
주최는 622화 + 3시간 전



3X3x3큐브공식 한번들으면 외워  
집 [개미버전]

더보기

[최근에 업로드된 동영상](#) [추천 동영상](#)



\*유출\* 슈퍼셀이랑 먼 겁니다..  
신규 전설 카드 '고스트' 유출합니  
피파북 클럽 ☺  
조회수 19 7만회 · 1일 전



기사 vs 왕 vs 고향 (성태 전쟁 미드)  
단호다무브스  
총회수 1.5천회 + 6시간 전



ND 미니  
프로수 4.3천원, 2시간 전



【범죄도시】 팔미궁에 손이 안닿는  
마동석

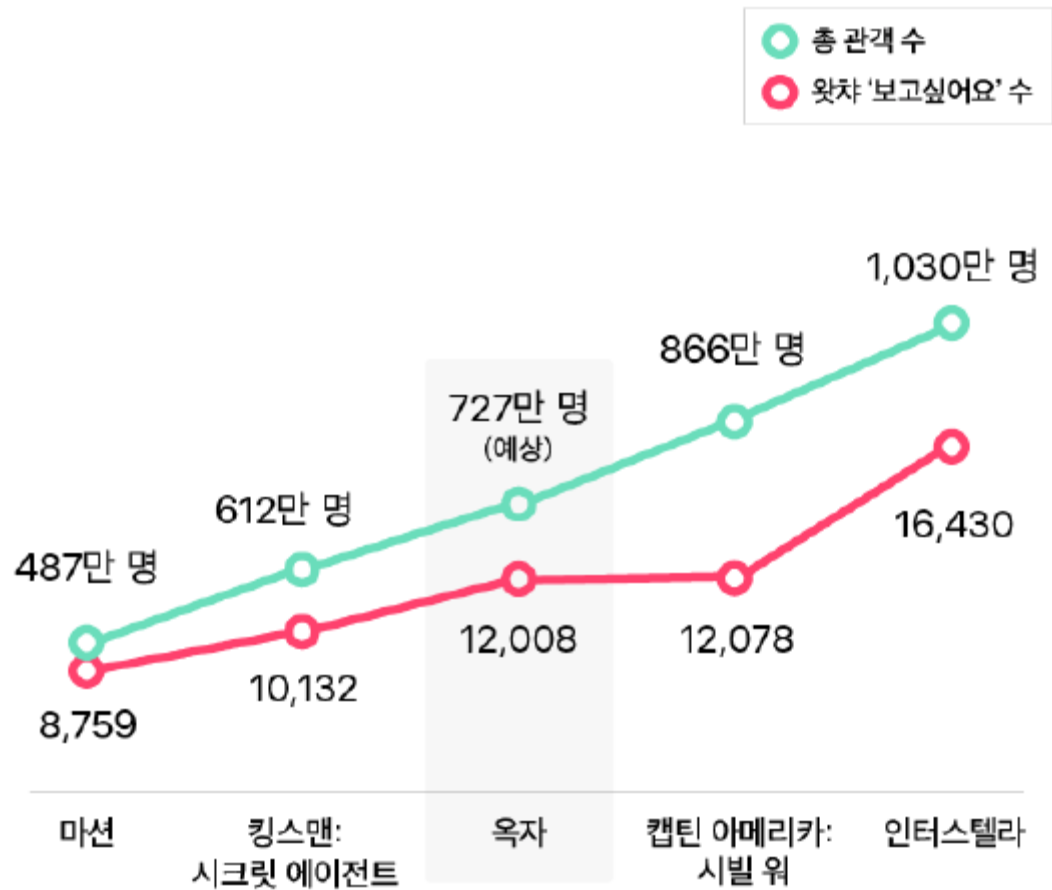
최익 TV  
주최수 952회 - 3시간 전



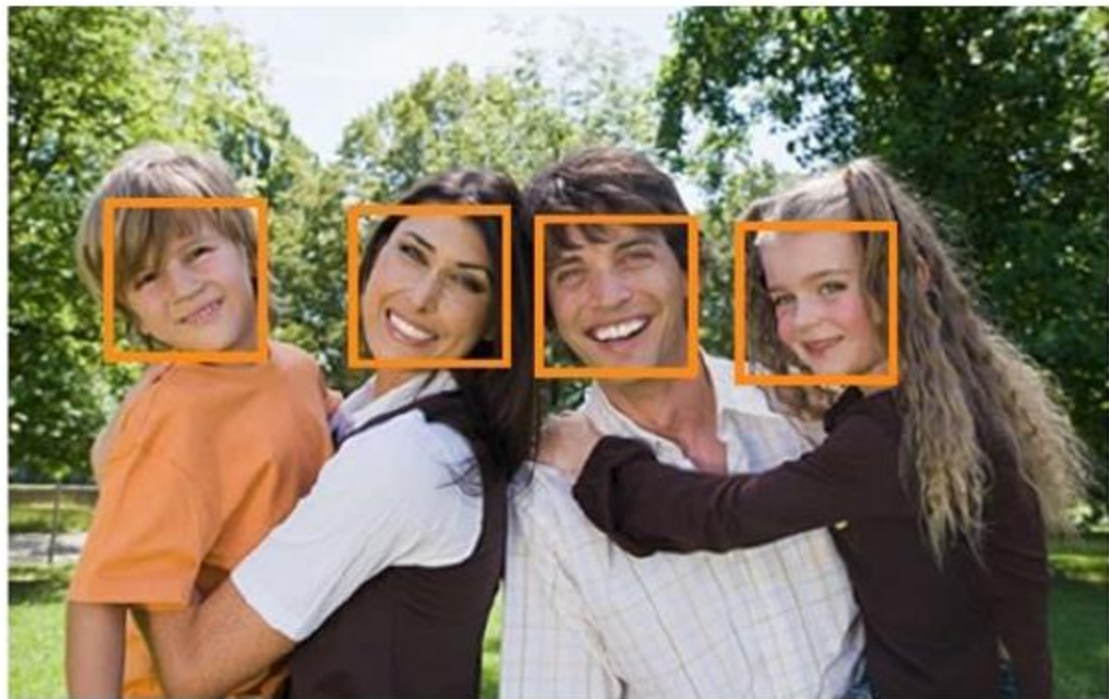
【별첨도시】 이동석한테 욕하는 위  
성락 작살나다  
초막 TV  
조별수 1.3천원 - 3시간 전



2505년, IQ 80이 미국 대통령이 될 수 있는 이유는?  
드림텔러(DreamTeller)  
주최수 7,2만회 · 5시간 전









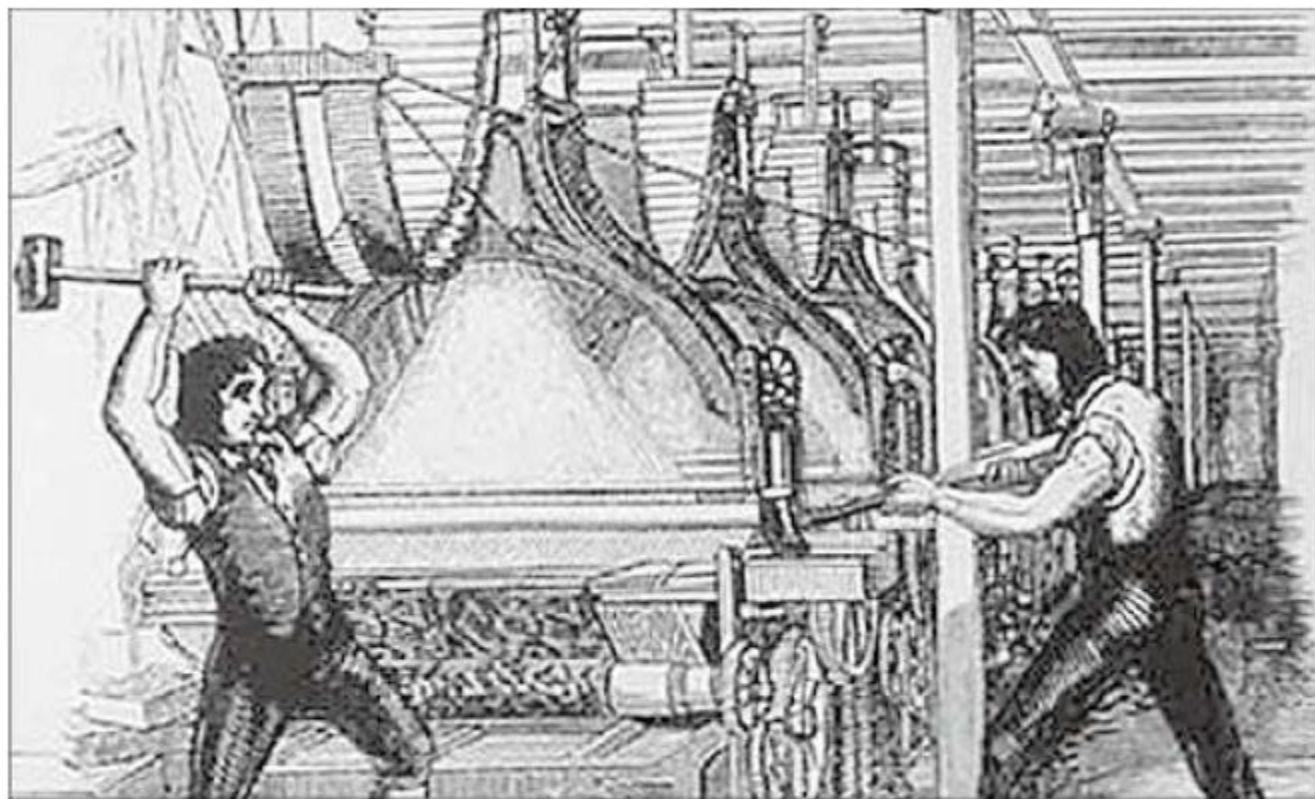






# 산업혁명

인간의노동 ➡ 증기기관



**사람 → 기계**

# 4차 산업혁명

**머신러닝**

**=**

**사람의 지적업무를 대신함**



# **머신러닝**

## **인간의지식 → 알고리즘**



# Machine Learning

산업혁명 - **물건** 생산을 위해 **육체노동**을 기계로 자동화하자

머신러닝 - **정보** 추출을 위해 **정신노동**을 기계로 자동화하자

(특히 deep learning)

## Key concepts

- Big Data
- 정해진 시간내에 **한 대의 머신**으로 처리가 어려운 크기의 데이터
  - 고성능 머신 또는 **다수의 머신**을 사용해 **목표**대로 데이터를 처리할 수 있음
  - ex) 실시간 대용량 SQL 처리

## Key concepts

- Machine Learning
- 데이터로 부터 패턴을 학습하는 기법  
→ 알고리즘, 모델의 묶음? 상위개념
  - 데이터의 크기는 중요,  
그러나 작다고 불가능한 것은 아님
  - ex) 컴퓨터에서 손 글씨 이해시키기

## Key concepts

### Data Mining

- 사실상 머신러닝과 같은 의미  
→ 데이터에서 의미있는 규칙 찾기
- 주요적용 사례가 비즈니스 분야
- 머신러닝 - 알고리즘  
데이터마이닝 - 어플리케이션
- ex) 스팸 필터, 상품 추천

# Key concepts

Big Data

활용

Data Mining

분산처리  
데이터베이스

기술개념

Machine  
Learning

Hadoop  
NoSQL Spark  
BigTable

알고리즘  
구현체

Navie Bayes  
SVM Logistics  
Regression

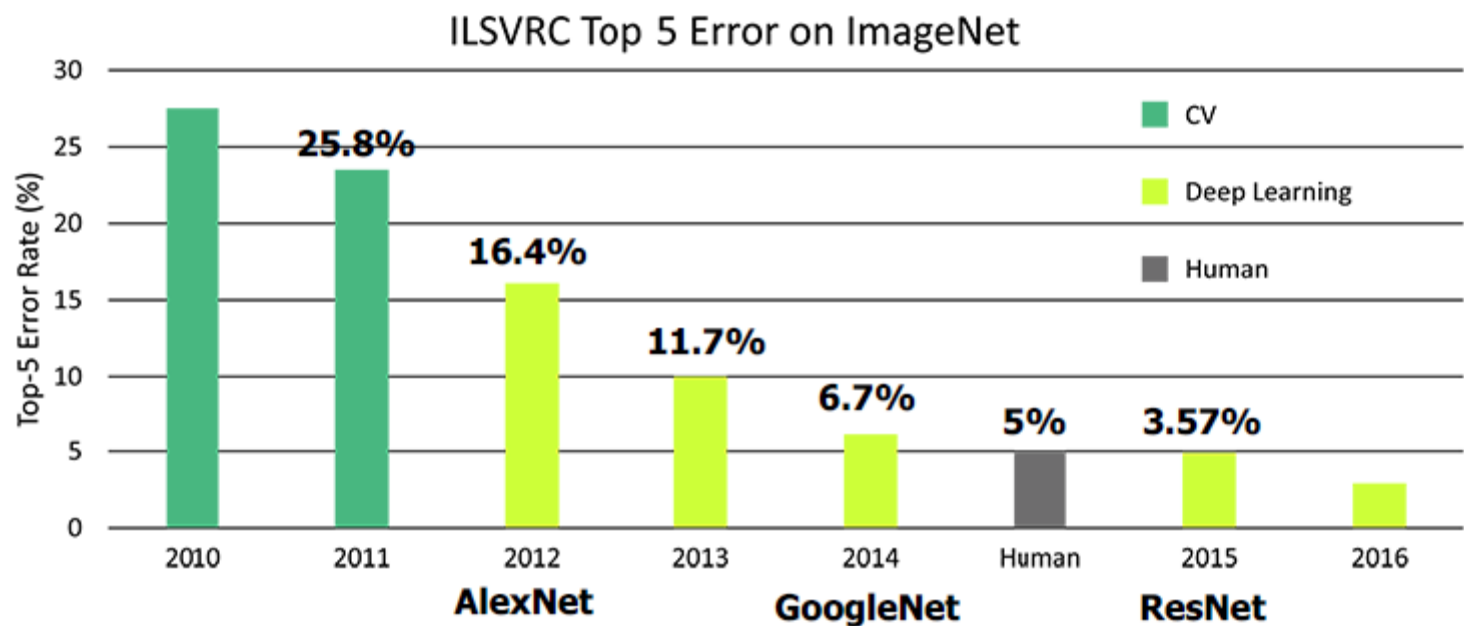


## Key concepts

### Deep Learning

- Neural Net 기반 머신러닝 알고리즘  
→ 최근 높은 성능에 대세로 떠오름
- 사진내 객체 인식, 번역 등 탁월한 성능
- 머신러닝과 구분하여, 새로운 분야로 주장하는 사람도 있음

# 인간보다 낮은 이미지 오인식률



Source - <https://goo.gl/UrBKWa>

# ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



# MACHINE LEARNING

Machine learning begins to flourish.



# DEEP LEARNING

Deep learning breakthroughs drive AI boom.



1950's

1960's

1970's

1980's

1990's

2000's

2010's

# Types of Machine Learning

**Supervised**

Y 데이터가 있는 것

**Regression**

**Classification**

**Unsupervised**

Y 데이터가 없는 것

**Clustering**

# Types of Machine Learning

회귀 – 데이터( $X, Y$ )를 줄께,  $X'$ 의 숫자  $Y'$ 를 예측해봐(Continuous  $Y$ )

ex) 주가 예측, 경제 성장률 예측, 영화 관람객 예측

분류 – 데이터( $X, Y$ )를 줄께,  $X'$ 의  $Y'$ 를 분류해봐(Discrete  $Y$ )

ex) 스팸메일 여부, 게임 어뷰저 여부, 신문기사 분류

군집 – 데이터( $X$ )를 줄께,  $X$ 를  $n$ 개로 나눠봐

ex) 고객 집단 나누기

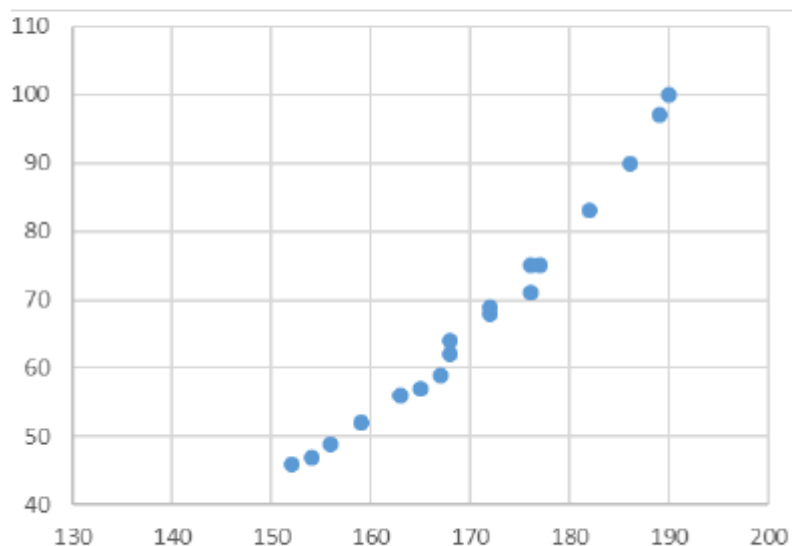
# Regression



## 키를 가지고 몸무게를 예측할 수 있을까?

Height	Weight	Height	Weight
152	46	172	69
154	47	172	68
156	49	176	71
159	52	176	75
163	56	177	75
165	57	182	83
167	59	186	90
168	64	189	97
168	62	190	100

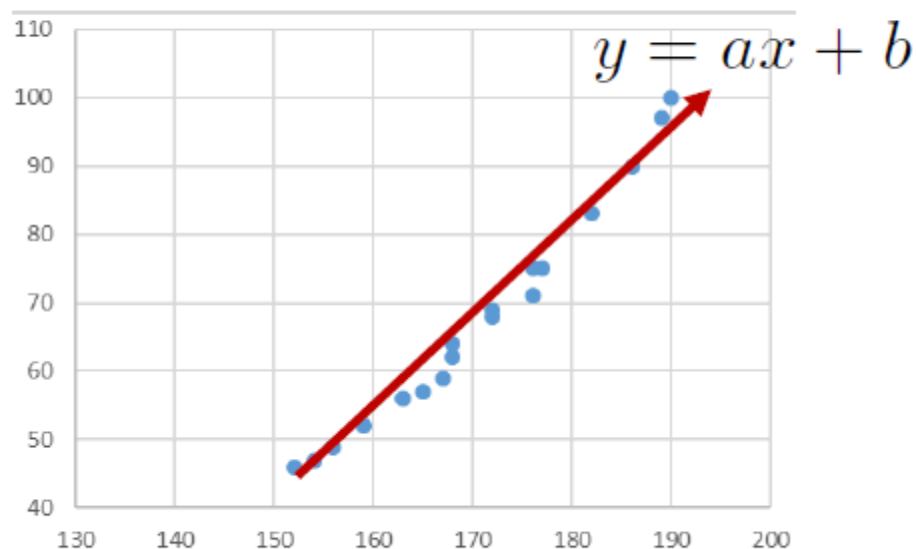
Data source: <http://goo.gl/gDscUQ>



# 키를 가지고 몸무게를 예측할 수 있을까?

Height	Weight	Height	Weight
152	46	172	69
154	47	172	68
156	49	176	71
159	52	176	75
163	56	177	75
165	57	182	83
167	59	186	90
168	64	189	97
168	62	190	100

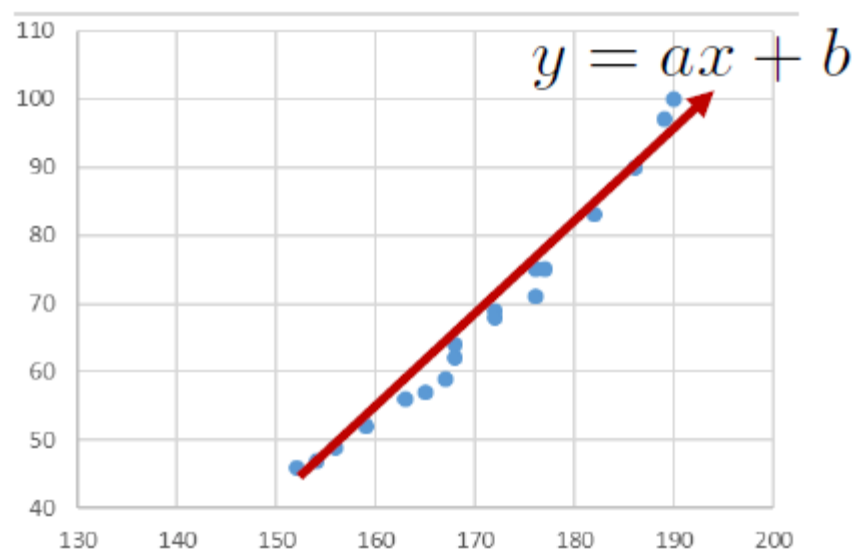
Data source: <http://goo.gl/gDscUQ>



## Regression – 회귀, 추세선을 긋는 것

Height	Weight	Height	Weight
152	46	172	69
154	47	172	68
156	49	176	71
159	52	176	75
163	56	177	75
165	57	182	83
167	59	186	90
168	64	189	97
168	62	190	100

Data source: <http://goo.gl/gDscUQ>

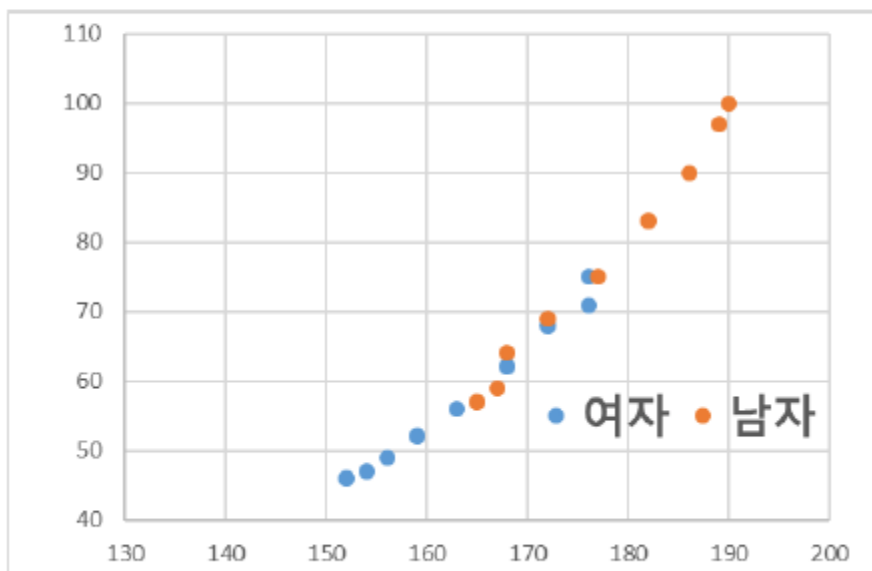


# Classification

# 키와 몸무게를 가지고 남/녀를 분류할 수 있을까?

H 186, W 82 남자? 여자?

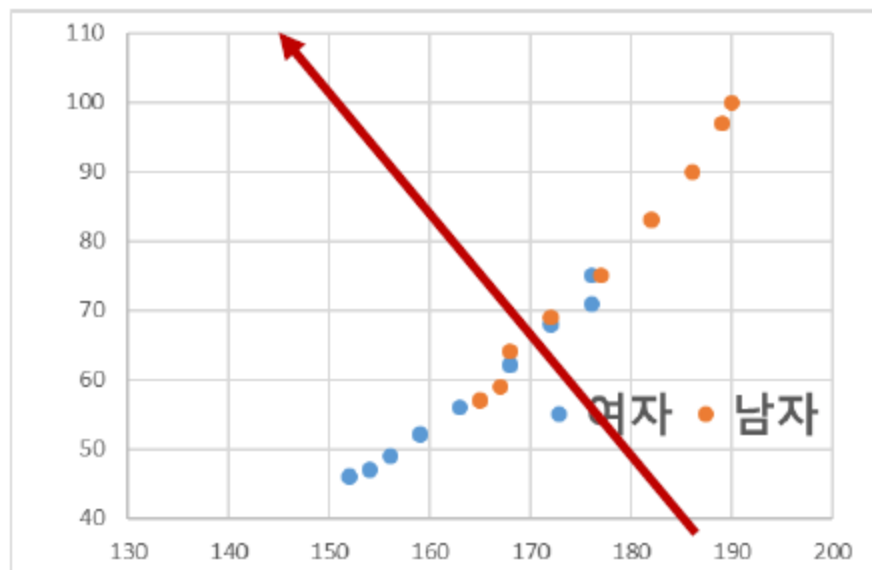
Sex	Height	Weight	Sex	Height	Weight
여자	152	46	남자	172	69
여자	154	47	여자	172	68
여자	156	49	여자	176	71
여자	159	52	여자	176	75
여자	163	56	남자	177	75
남자	165	57	남자	182	83
남자	167	59	남자	186	90
남자	168	64	남자	189	97
여자	168	62	남자	190	100



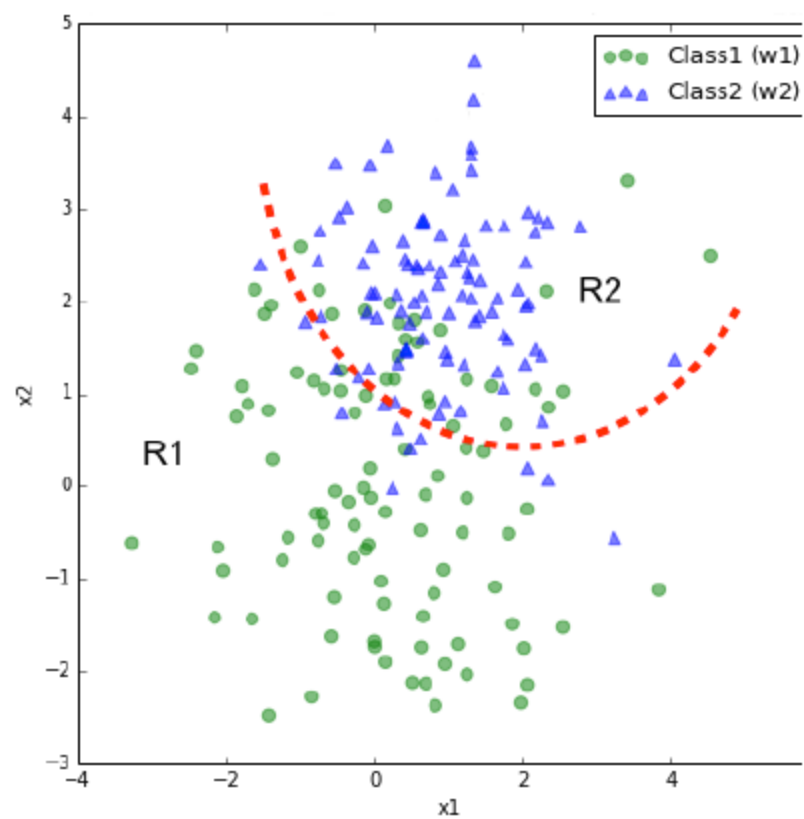
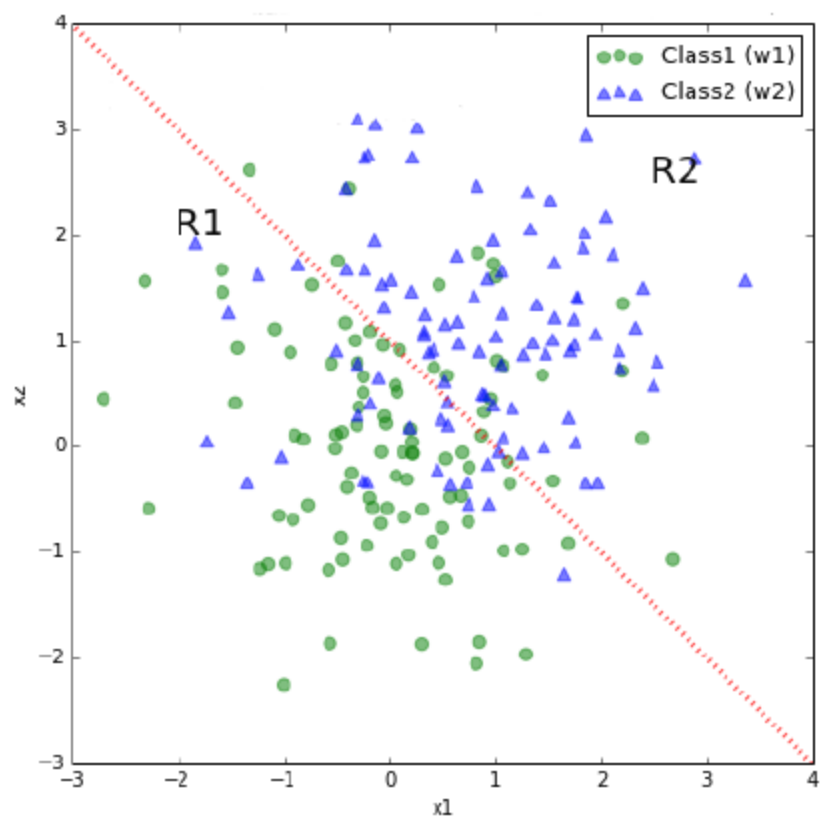
# Classification – 분류, 데이터의 유형을 나누는 것

기존 데이터를 바탕으로 데이터 유형을 나눠 보기

Sex	Height	Weight	Sex	Height	Weight
여자	152	46	남자	172	69
여자	154	47	여자	172	68
여자	156	49	여자	176	71
여자	159	52	여자	176	75
여자	163	56	남자	177	75
남자	165	57	남자	182	83
남자	167	59	남자	186	90
남자	168	64	남자	189	97
여자	168	62	남자	190	100

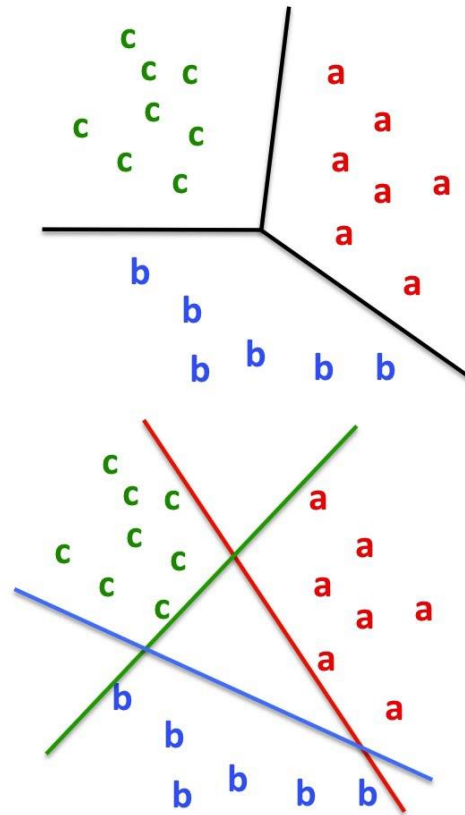






# Multi-class vs. Binary classification

- Multi-class:
  - classes mutually exclusive:
    - instance is either a or b or c
    - even if it's an outlier
  - NB, kNN, DT, logistic
- Binary:
  - one-vs-rest:
    - {a} vs {not a}, {b} vs {not b}
  - classes may overlap
    - instance can be both a and b
    - can be in none of the classes
  - SVM, logistic, perceptron

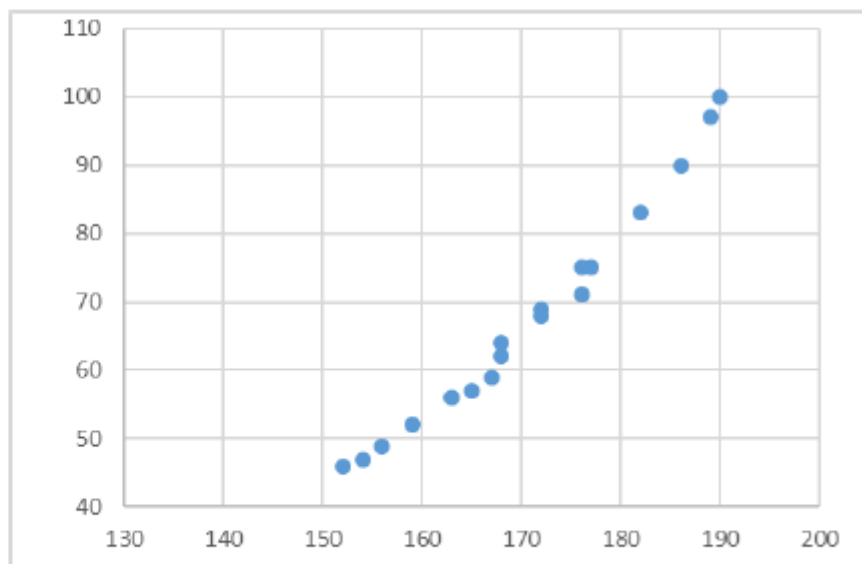


# Clustering

# 농구팀을 만든다면 몇 팀으로 만들까?

어떻게 사람을 묶으면 될까?

Height	Weight	Height	Weight
152	46	172	69
154	47	172	68
156	49	176	71
159	52	176	75
163	56	177	75
165	57	182	83
167	59	186	90
168	64	189	97
168	62	190	100



# Clustering – 군집, 데이터를 모으는 것

## 아무런 사전정보 없이 데이터 유형을 나눠보기

Sex	Height	Weight	Sex	Height	Weight
여자	152	46	남자	172	69
여자	154	47	여자	172	68
여자	156	49	여자	176	71
여자	159	52	여자	176	75
여자	163	56	남자	177	75
남자	165	57	남자	182	83
남자	167	59	남자	186	90
남자	168	64	남자	189	97
여자	168	62	남자	190	100

