

인공지능과 머신러닝

1. 예측 (Prediction)

제주대학교 변영철



나, 연서





양, 흙끼



한국





한국
서예
미술
도







나만의 메리풀드 ⁸⁰³

날짜	길이(전체)	너비	길이(앞다리)
2019.06.16	0	0	0
2019.06.20	1.6cm	0.28cm	0.9cm
06.22	2.6cm	3.8cm	1.4cm
6.25	3cm	4.6cm	1.6cm
2019.6.29	4.4 5.0	1.8cm 0.5cm	2cm 2.2

① 쿠다란

2019.06.16	0	0	0
2019.06.20	1.6cm	0.28cm	0.9cm
2019.06.22	3.3cm	4.5cm	1.5cm
6.25	3.5cm	5cm	1.8cm
6.29.	5cm	0.5cm	2.5cm

2

돌진

3

돌진

날짜	길이	너비	길이
6.16	0	0	8mm
6.20	12mm	2.9mm	8mm
6.22	25mm	4.5mm	13mm
25	31m	5.5mm	15mm
6.29	45.5mm	60mm	19mm

지난날짜, 키, 잎 너비, 잎 길이, 주인

1, 0, 0, 0, 1

1, 0, 0, 0, 2

1, 0, 0, 0, 3

5, 16, 28, 9, 1

5, 16, 2.8, 9, 2

5, 12, 2.9, 8, 3

7, 33, 4.5, 15, 1

7, 26, 3.8, 14, 2

7, 25, 4.5, 13, 3

10, 35, 5, 18, 1

10, 30, 4.6, 16, 2

10, 31, 5.5, 15, 3

14, 50, 6.5, 25, 1

14, 44, 5.8, 20, 2

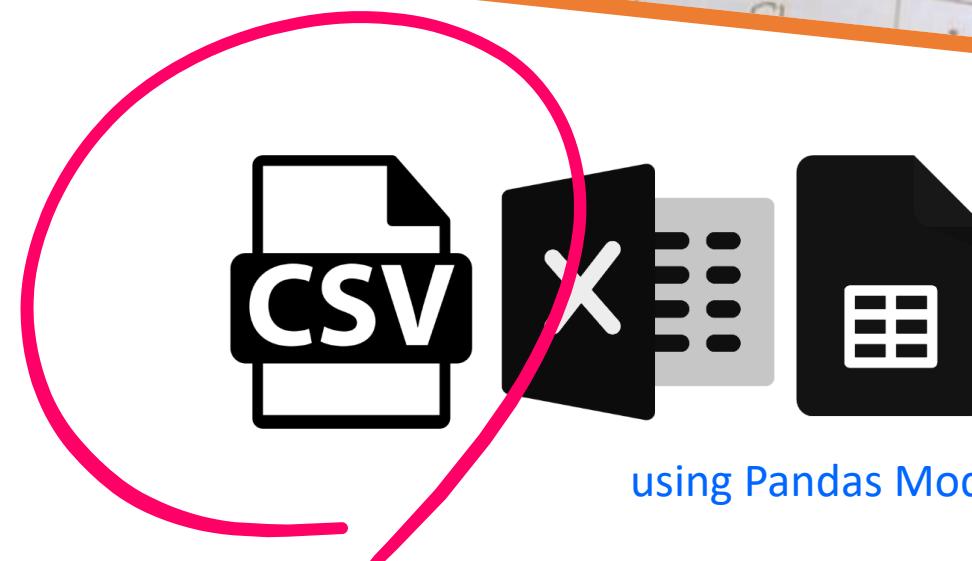
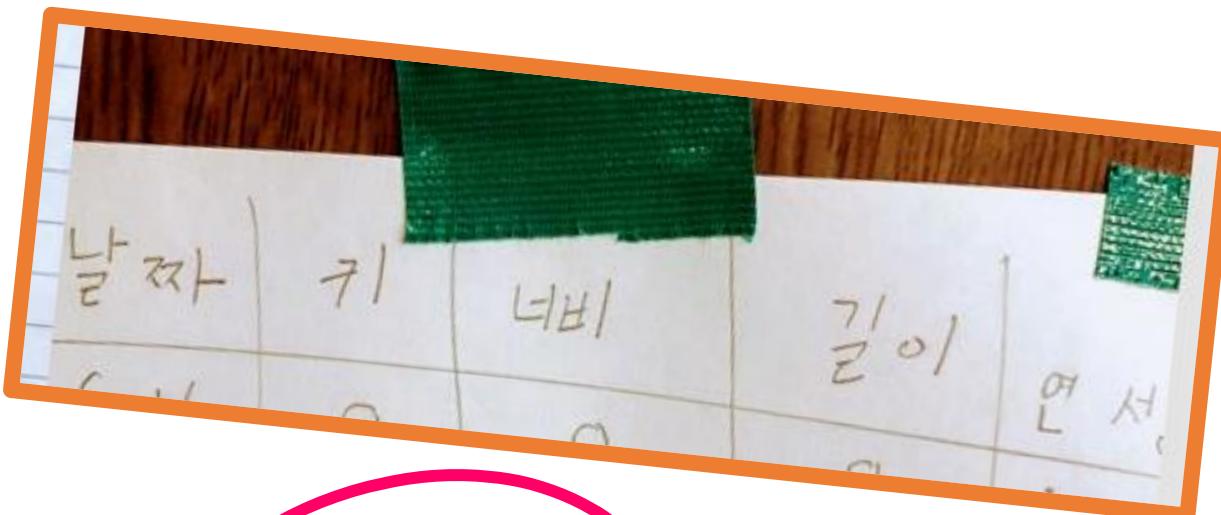
14, 45.5, 6, 19, 3

20, 56, 6.8, 27, 1

20, 50, 6, 22, 2

20, 51, 6.5, 21, 3

plant_diary.csv



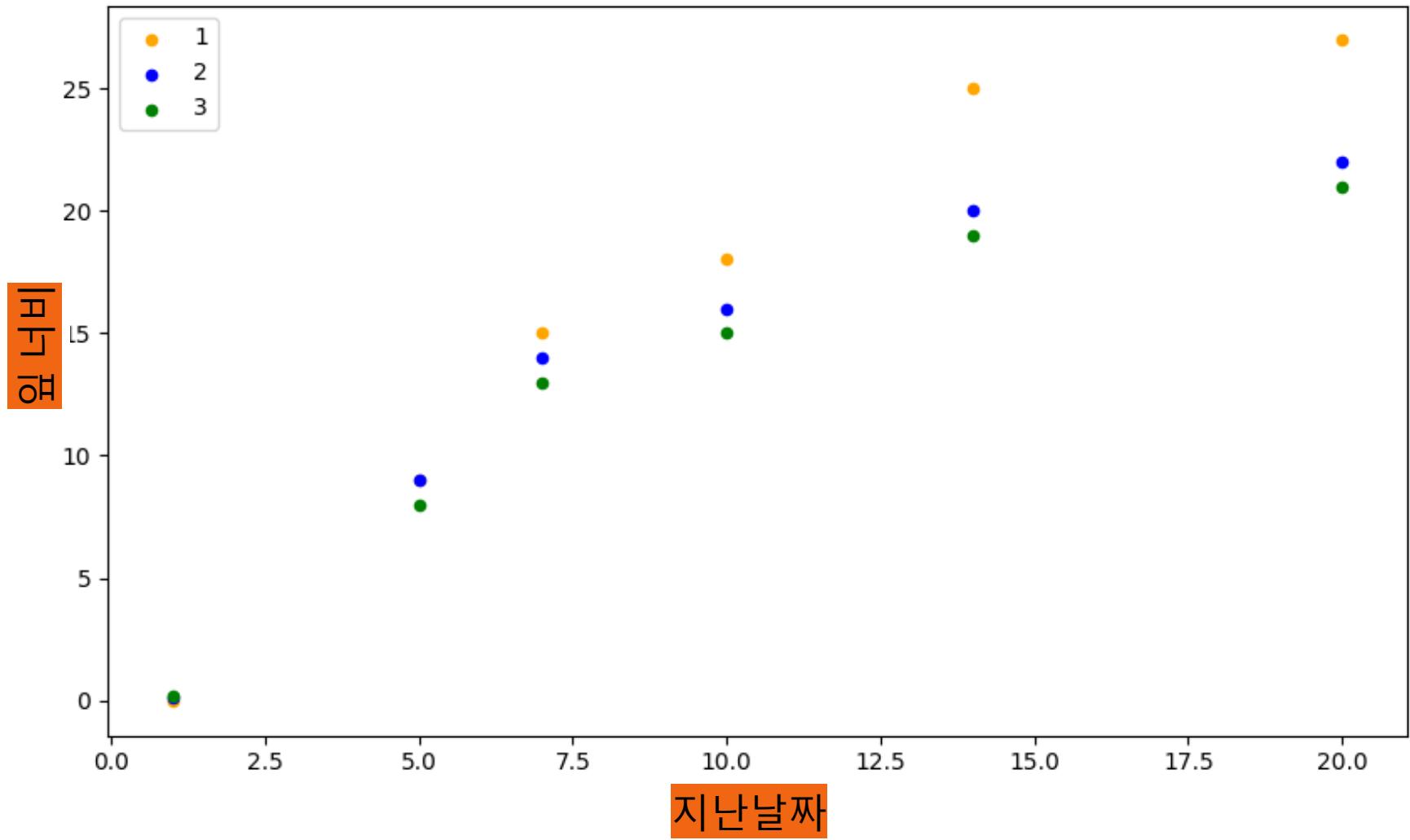
using Pandas Module

'날짜'가 지남에 따라
'잎 너비'는 얼마나 자랐을까?
점으로 찍어봐라(**plot**)!
(주인에 따라 다른 색으로 표시)

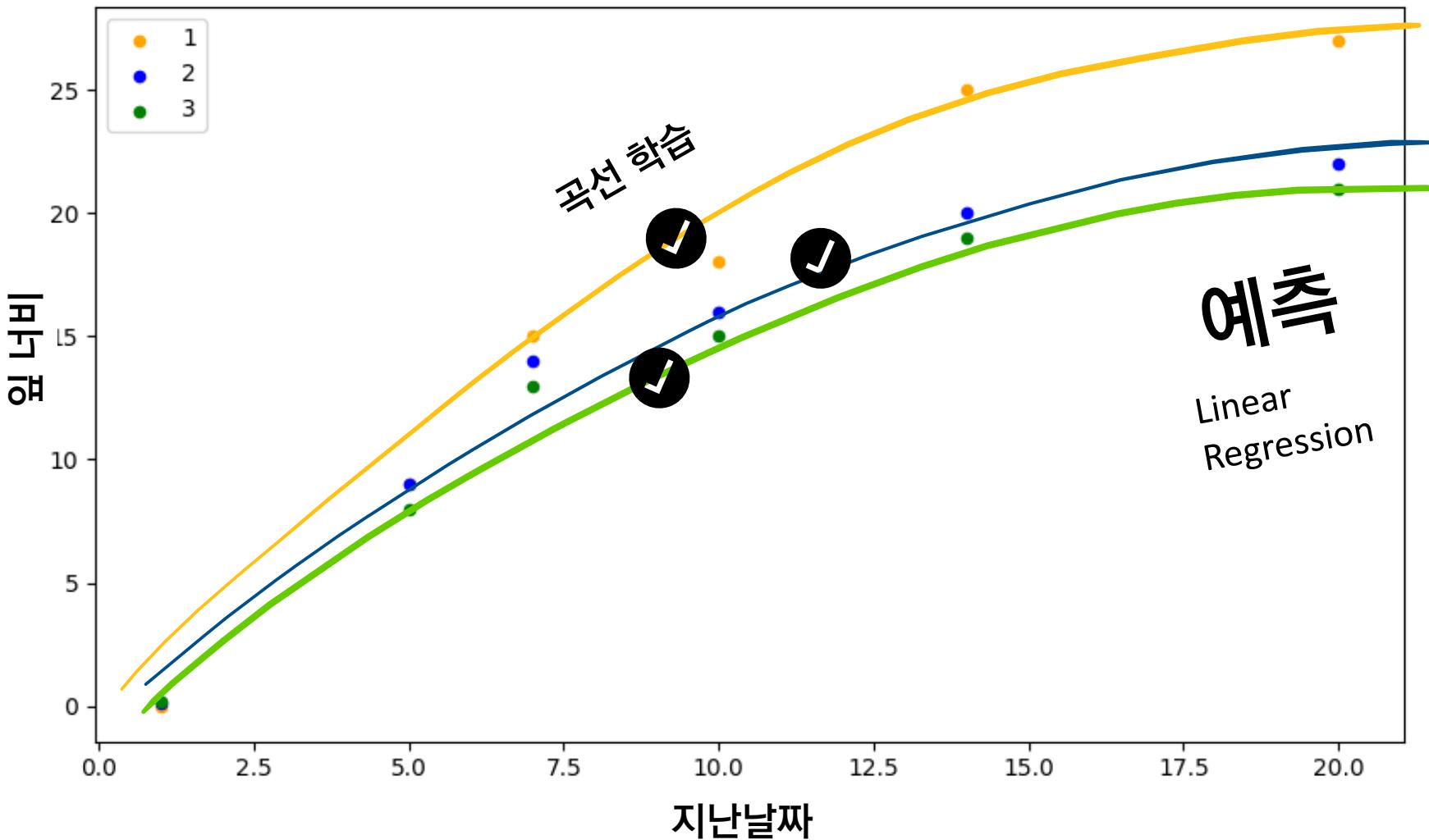
plot(df, '날짜', '잎 너비', '주인')

아까 그 데이터

첫째 아이, 둘째
셋째



지난날짜 vs. 잎 너비



67,078,401.33 KRW

+66,169,933.82 (7,283.70%) ↑ 지난 5년

12월 2일 오전 12:35 UTC · 면책조항

1일 | 5일 | 1개월 | 6개월 | 연증 | 1년 | 5년 | 최대



1

BTC ▾

67078401.33

KRW ▾

67,078,401.33 KRW

+45,946,042.26 (217.42%) ↑ 지난해

12월 2일 오전 12:35 UTC · 면책조항

1일 | 5일 | 1개월 | 6개월 | 연중 | 1년 | 5년 | 최대



1

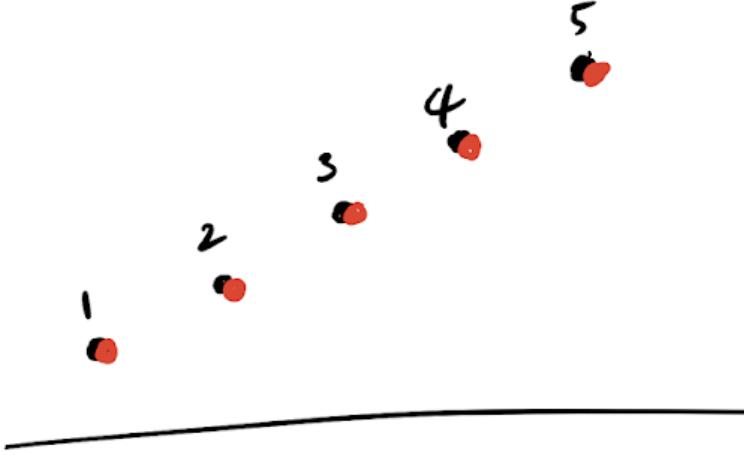
BTC ▾

67078401.33

KRW ▾

스코어(score)

- 정답
- 대답한 값(예측값)



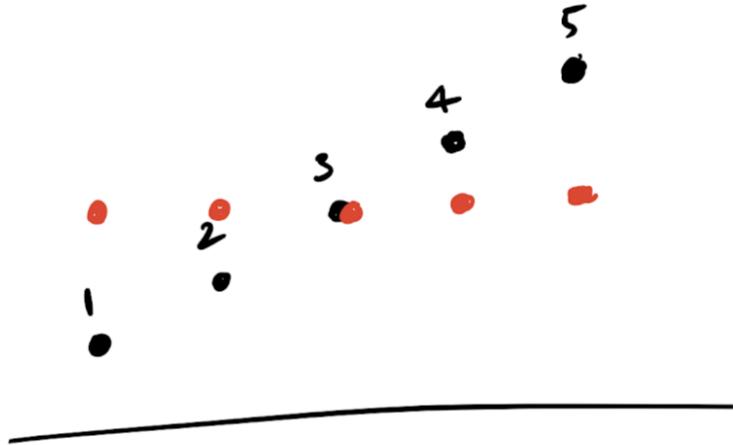
$$\text{Score} = 1 - \frac{(\bullet - \bullet)^2}{(\bullet - \bar{x})^2} = 1 - \frac{0}{10} = 1$$

$\sim\sim\sim$

$$\begin{aligned} & (1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 \\ &= 2^2 + 1^2 + 0^2 + 1^2 + 2^2 = 10 \end{aligned}$$

스코어(score)

- 정답
- 대답한 값(예측값)



$$\text{Score} = 1 - \frac{(\bullet - \bullet)^2}{(\bullet - \bar{x}_{\text{정답}})^2} = 1 - \frac{10}{10} = 1 - 1 = 0$$

$$\begin{aligned}&= (1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 \\&= 2^2 + 1^2 + 0^2 + 1^2 + 2^2 = 10\end{aligned}$$

예측 알고리즘

Machine Learning

- KNeighborsRegressor (K-근접)
- DecisionTreeRegressor (결정 트리)
- RandomForestRegressor (랜덤 포레스트)
- Linear Regressor (선형 회귀)
- GradientBoostingRegressor (부스팅)
- XGBRegressor (부스팅)
- CatBoostRegressor (부스팅)

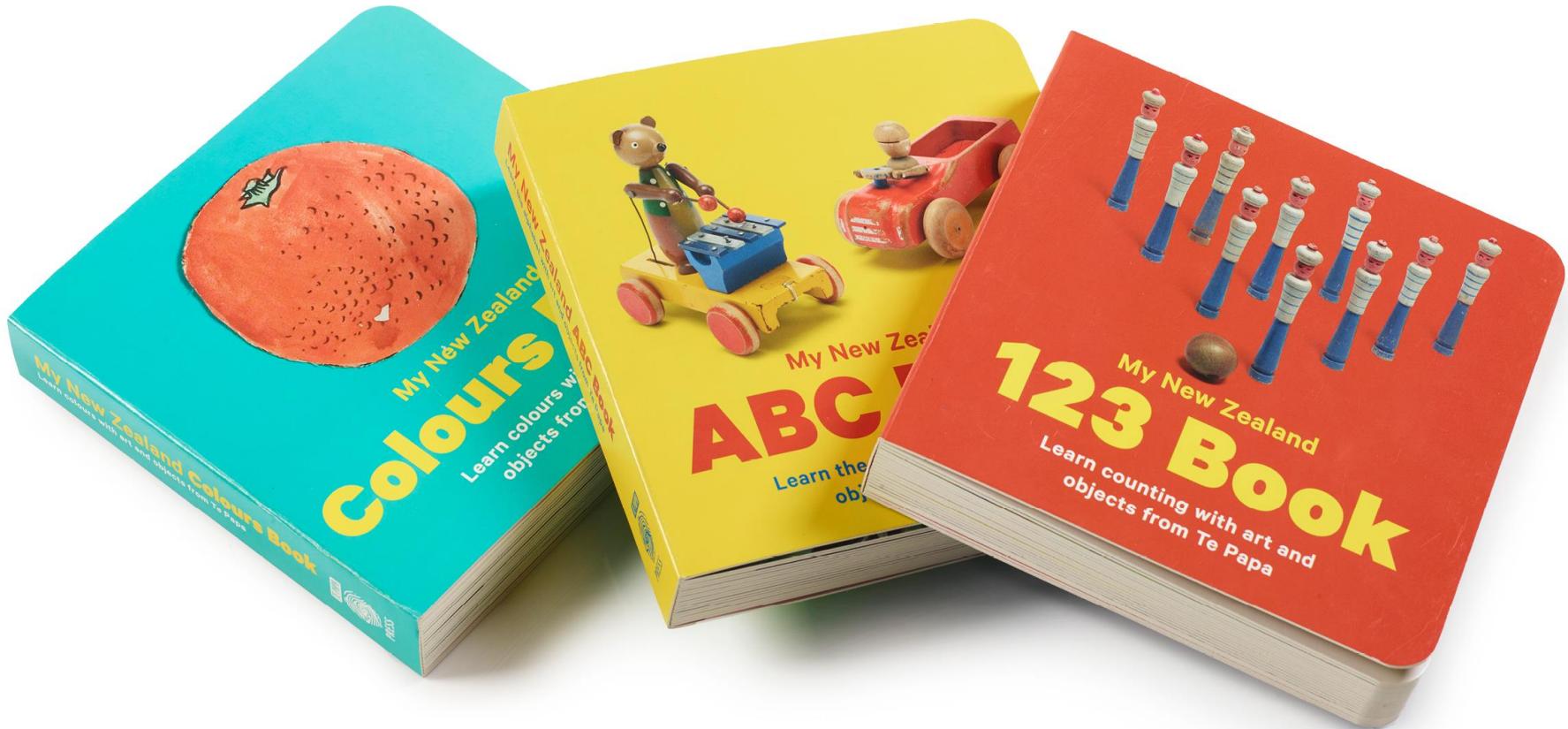
- NN-based LinearRegression
- MLPRegressor
- RNN/LSTM/GRU

Deep Learning



인공지능과 머신러닝

2. 분류 (Classification)

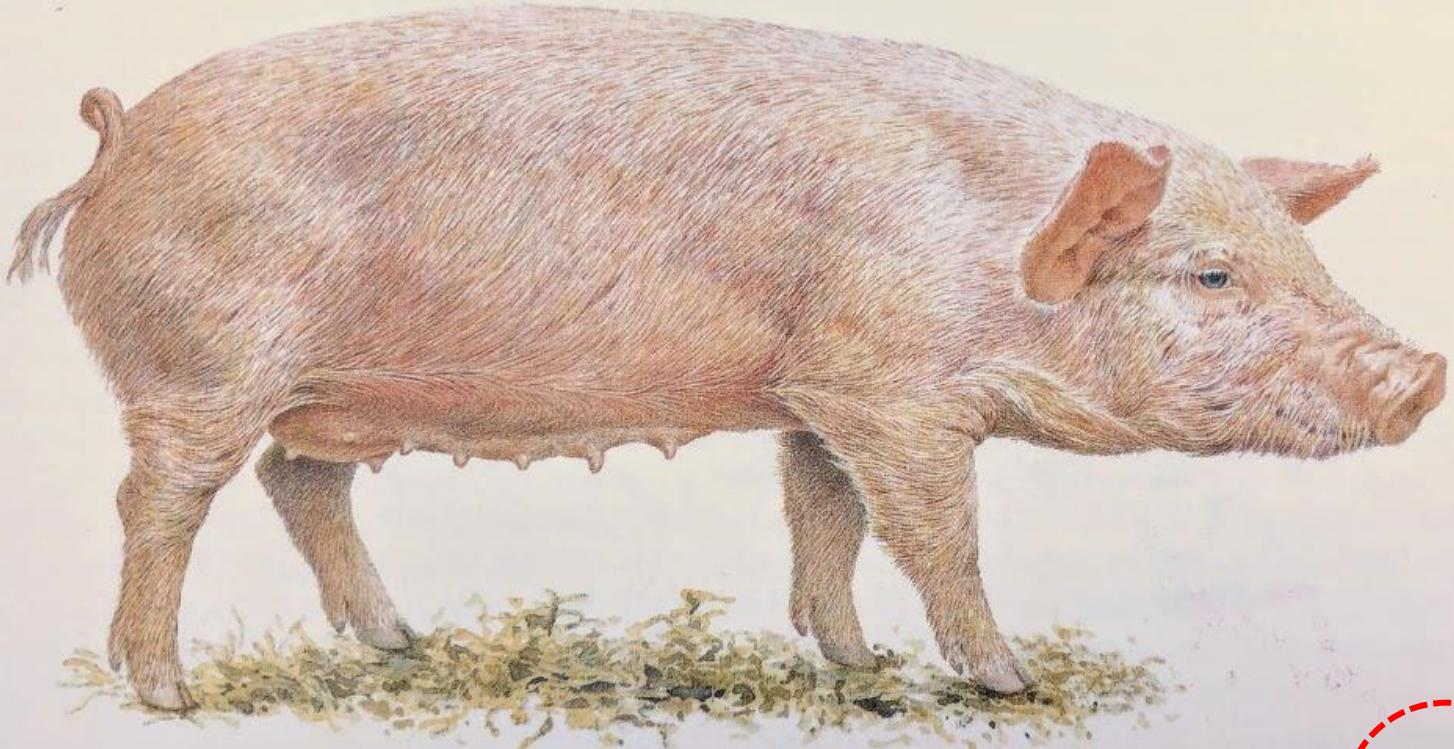




개



토끼



돼지



오리



지도학습

Supervised Learning

7

$$7 \times 1 = 7$$

$$7 \times 2 = 14$$

$$7 \times 3 = 21$$

$$7 \times 4 = 28$$

$$7 \times 5 = 35$$

$$7 \times 6 = 42$$

$$7 \times 7 = 49$$

$$7 \times 8 = 56$$

$$7 \times 9 = 63$$

첫째값, 둘째값, 정답

7, 1, 7

7, 2, 14

7, 3, 21

7, 4, 28

7, 5, 35

7, 6, 42

7, 7, 49

7, 8, 56

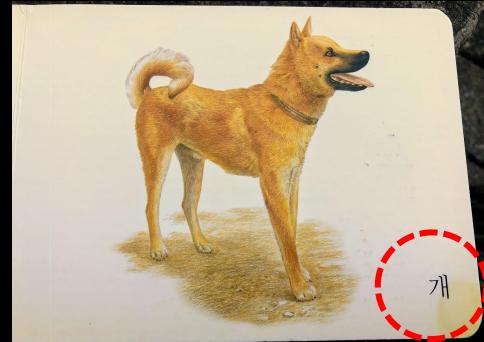
7, 9, 63

키
몸무게
발 크기
학년
성별



번호, 키, 몸무게, 발크기, 학년, 성별

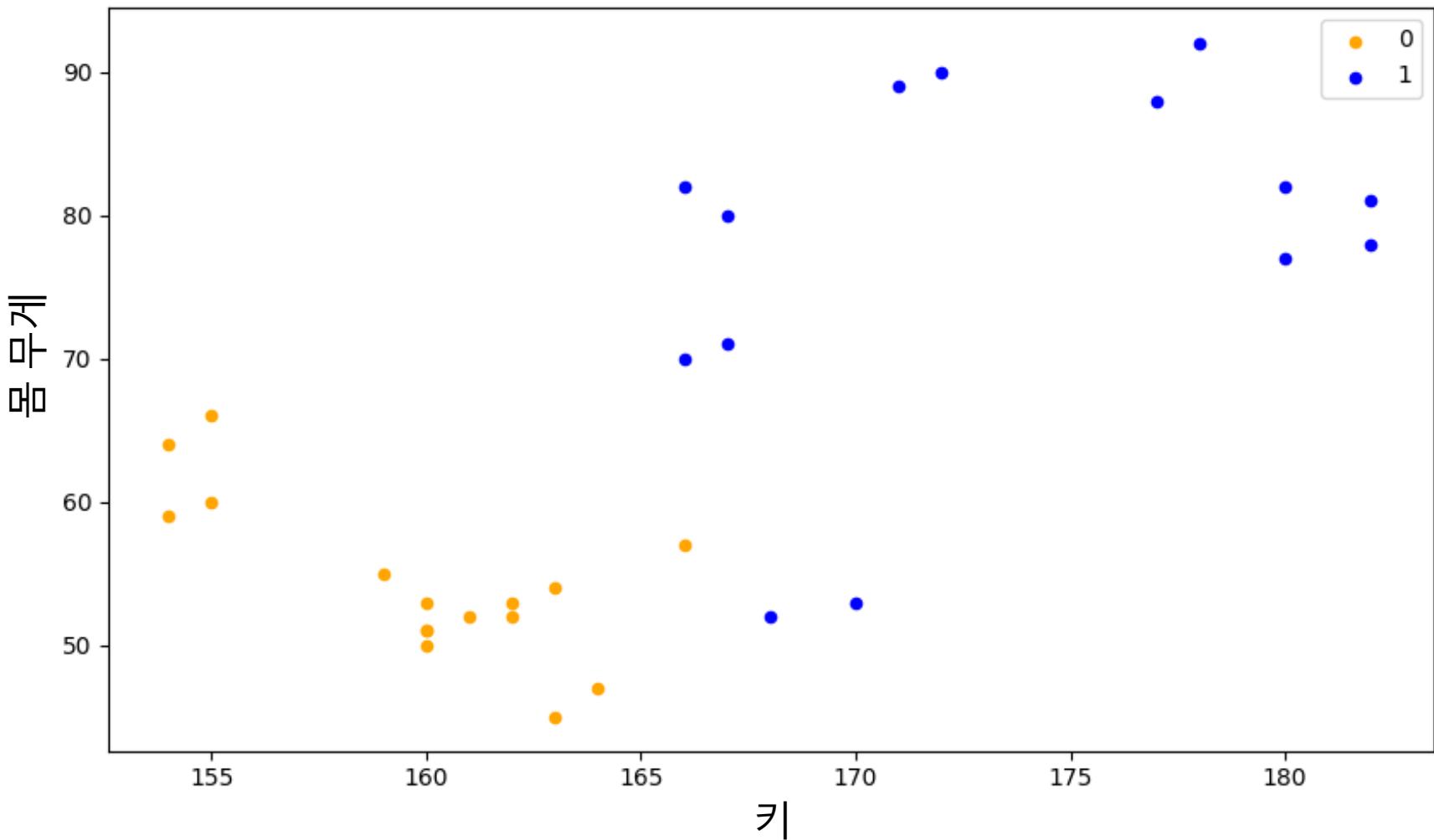
1,166,57,240,1,0
2,178,92,265,1,1
3,167,80,270,1,1
4,168,52,245,2,1
5,155,60,235,2,0
6,163,45,230,2,0
7,160,53,235,3,0
8,180,77,260,4,1
9,167,71,260,2,1
10,160,51,245,2,0
11,162,53,240,2,0
12,180,82,280,6,1
13,172,90,255,6,1
14,160,51,245,5,0
15,155,66,245,5,0
16,163,54,242,5,0
17,177,88,263,5,1
18,166,82,268,6,1
19,170,53,247,6,1
20,154,59,234,1,0
21,164,47,232,1,0



키에 따라
몸무게는 어떻게 변할까?
(성별에 따라 다른 색으로 표시)

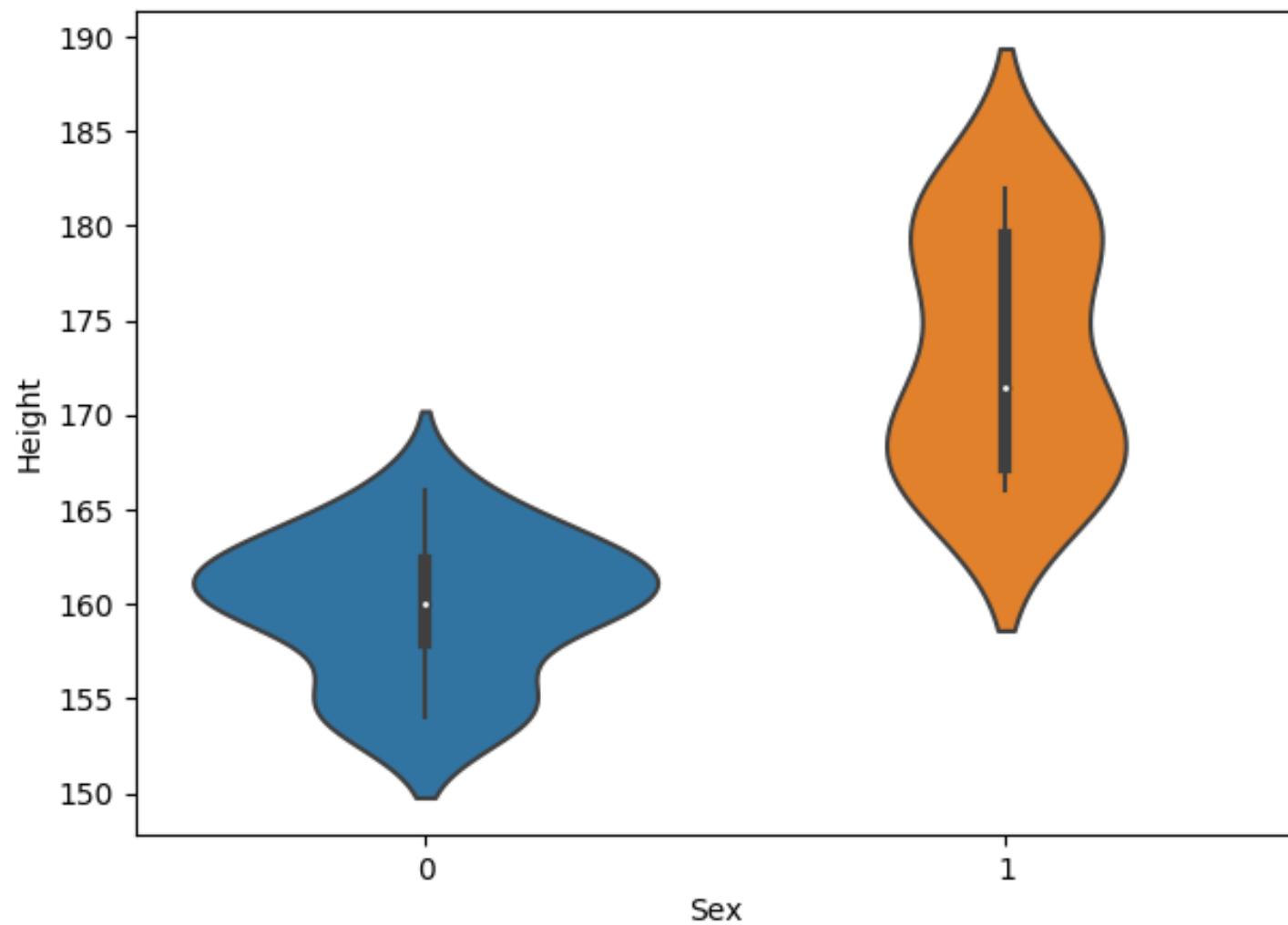
plot(df, '키', '몸무게', '성별')

$\text{키} |$ vs. 몸무게

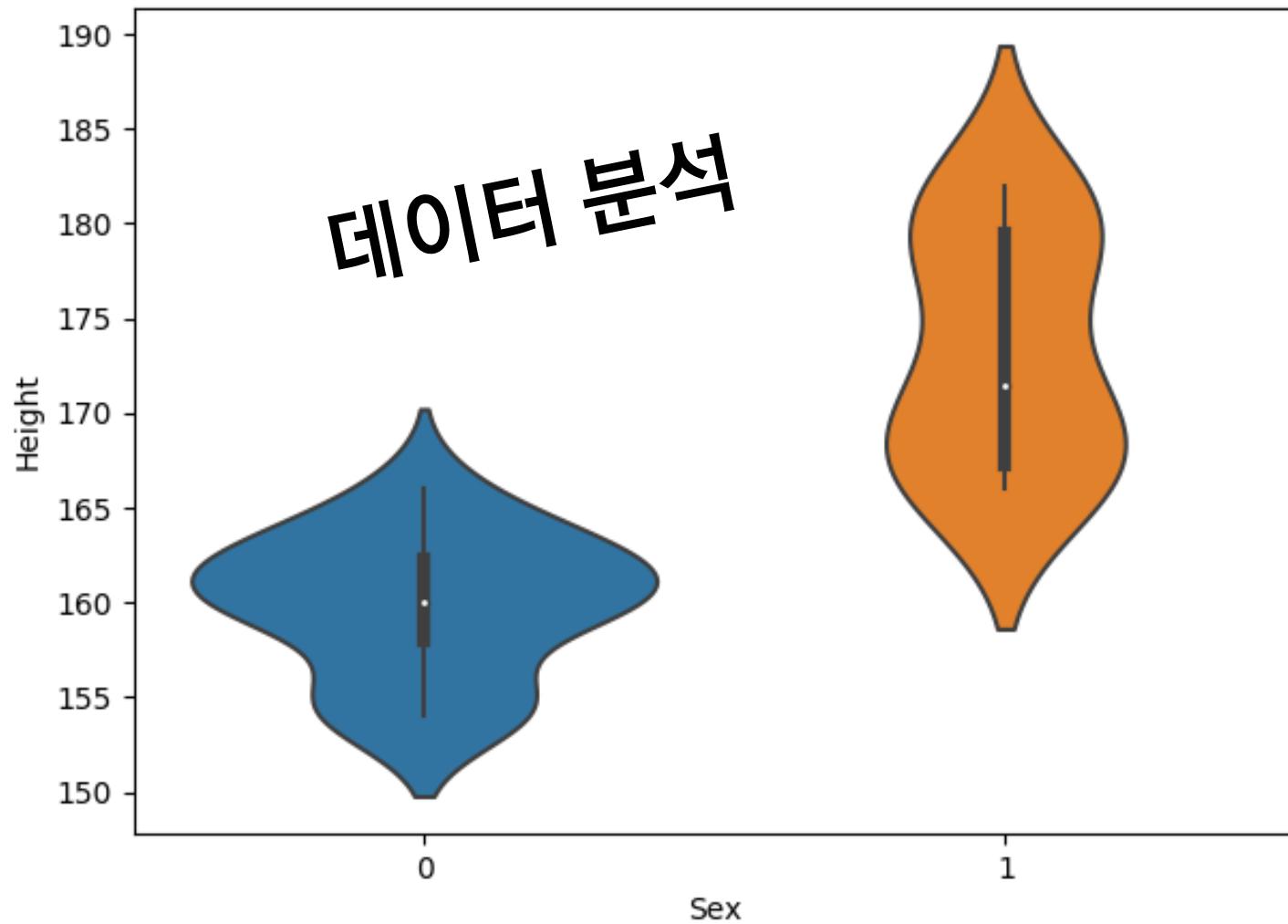


```
violinplot(df, '성별', '키')
```

성별에 따라 키가 어떻게 변하는지
바이올린 모양으로 표시해보라!



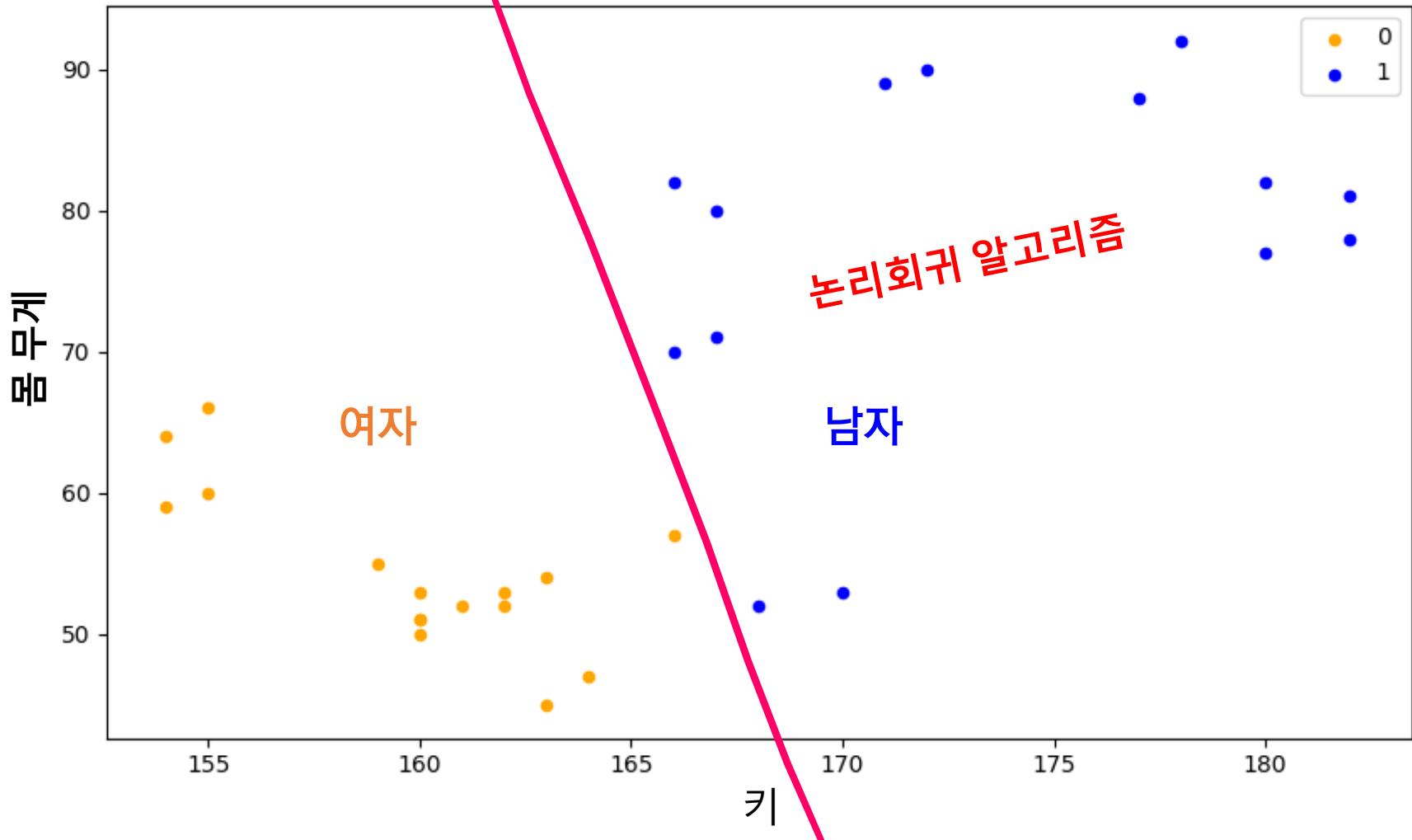
데이터 분석



분류

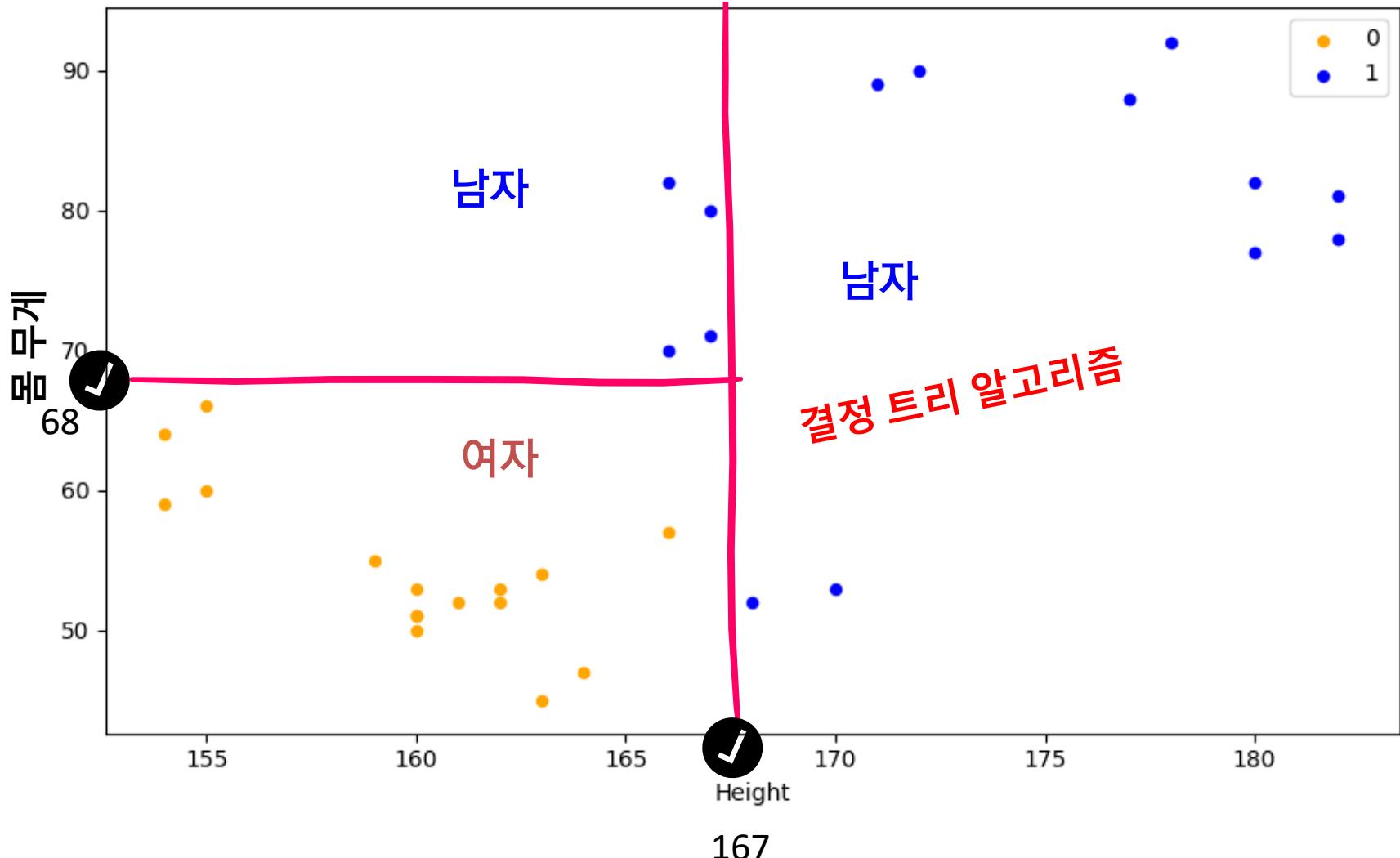
Classification

키 vs. 몸무게



분류

Classification 키 vs. 몸무게



키, 몸무게, 발 크기, 학년, 성별



학습용
문제

166,57,240,1,	0
178,92,265,1,	1
167,80,270,1,	1
168,52,245,2,	1
155,60,235,2,	0
163,45,230,2,	0
160,53,235,3,	0
180,77,260,4,	1
167,71,260,2,	1
160,51,245,2,	0
162,53,240,2,	0
180,82,280,6,	1
172,90,255,6,	1
160,51,245,5,	0

정답

테스트용
문제

155,66,245,5,	0
163,54,242,5,	0
177,88,263,5,	1
166,82,268,6,	1
170,53,247,6,	1
154,59,234,1,	0
164,47,232,1,	0

정답

키, 몸무게, 발 크기, 학년, 성별

학습용
문제

166,57,240,1,	0
178,92,265,1,	1
167,80,270,1,	1
168,52,245,2,	1
155,60,235,2,	0
163,45,230,2,	0
160,53,235,3,	0
180,77,260,4,	1
167,71,260,2,	1
160,51,245,2,	0
162,53,240,2,	0
180,82,280,6,	1
172,90,255,6,	1
160,51,245,5,	0

정답

테스트용
문제

155,66,245,5,	0
163,54,242,5,	0
177,88,263,5,	1
166,82,268,6,	1
170,53,247,6,	1
154,59,234,1,	0
164,47,232,1,	0

정답

gildong = SVC()

gildong.fit('학습용문제', '정답')

prediction= gildong.predict('테스트용문제')



분류 알고리즘

Machine Learning

- SVC (서포트벡터머신)
- DecisionTreeClassifier (결정트리)
- RandomForestClassifier (랜덤포레스트)
- XGBClassifier (XGBoost, eXtreme Gradient Boosting, Boosting or Additive Training)
(부스팅)
- LogisticRegression (논리회귀)
- Multilayer Neural Networks
- CNN/RCNN/GCNN

Deep Learning



지능을 갖다

(지능, *intelligence*, 知能)

새로운 사물 현상에 부딪쳐 지식을 활용하여 그 의미를 이해하고 처리 방법을 알아내는 지적 활동 능력

인공지능 (AI, Artificial Intelligence),
사람에 의한, 컴퓨터에 구현된 지능

인공지능

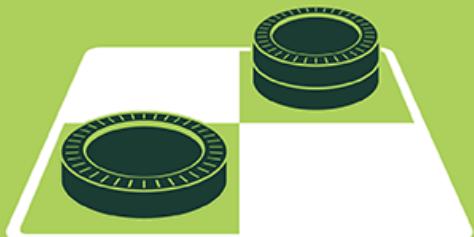
AI (Artificial Intelligence),
사람의 지능을 컴퓨터에
구현한 지능

머신러닝(기계학습)

입력과 정답(데이터, 경험, experience)을
주고 학습시켜 → 지식(knowledge)을
얻는 것.

ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



1950's

1960's

1970's

1980's

1990's

2000's

2010's

MACHINE LEARNING

Machine learning begins to flourish.



DEEP LEARNING

Deep learning breakthroughs drive AI boom.

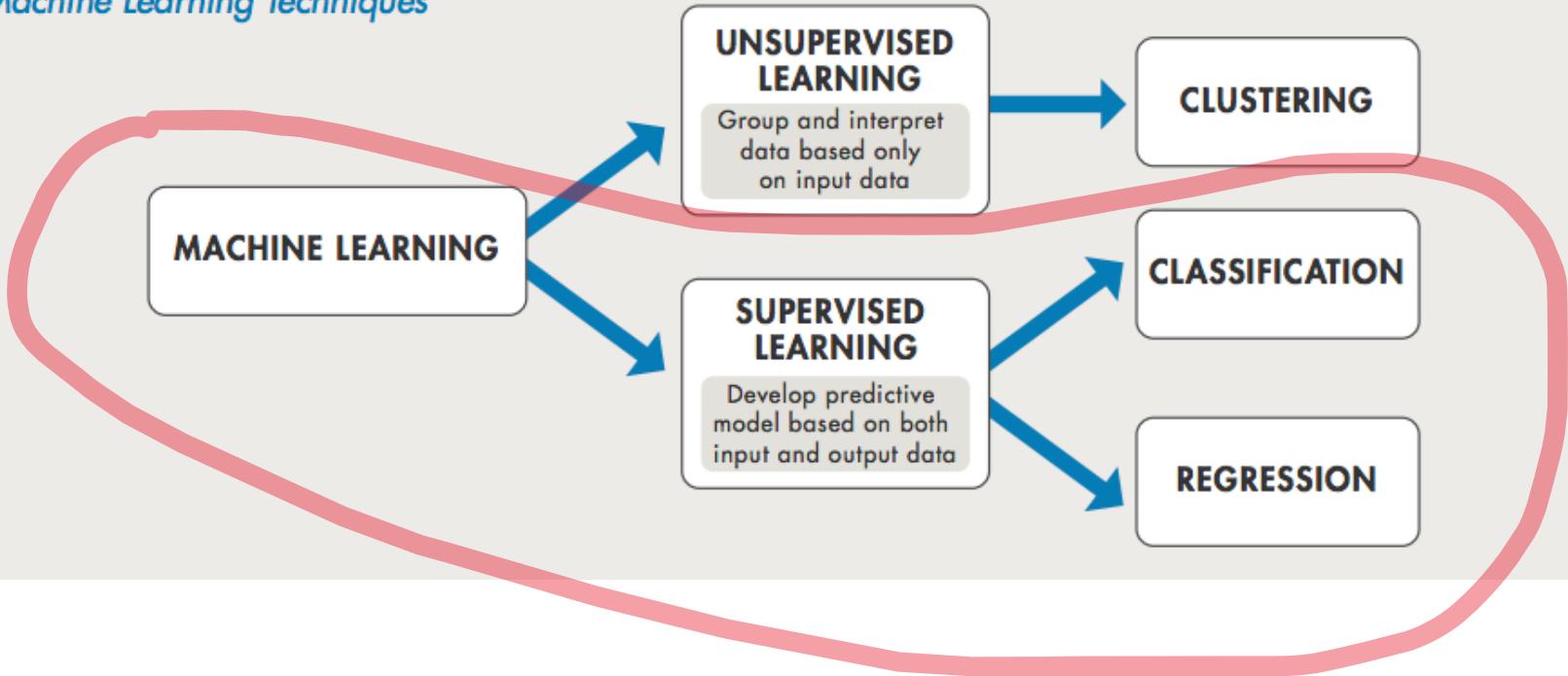


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

머신러닝 절차

1. CSV 파일 로드
2. 데이터 시각화와 분석
3. 데이터 전처리 (정규화, 데이터 균형)
4. 특징 상관관계 분석
5. 데이터 분할
6. ML 알고리즘을 이용한 분류/예측
7. 평가 및 결론

Machine Learning Techniques



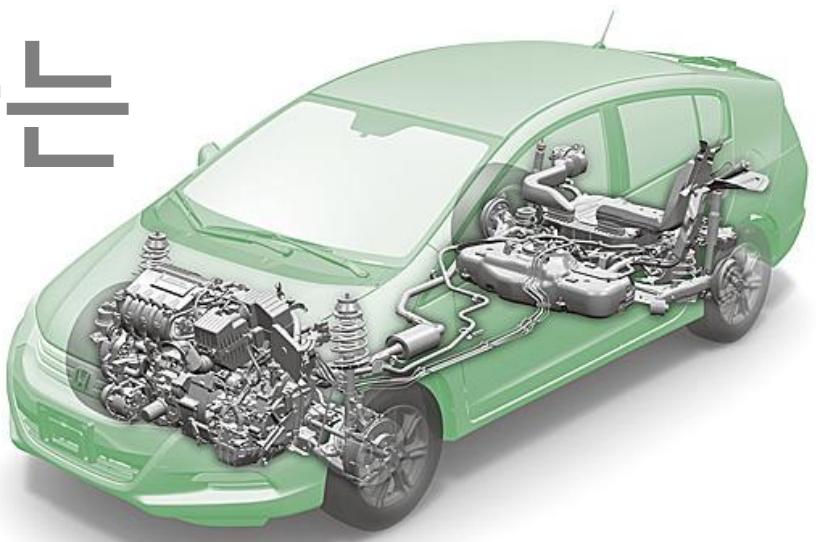
어떻게 배울까?

A close-up photograph of a man's face. He has a worried expression, with his hands resting against his temples and eyes looking directly at the viewer. A large red circle is drawn around the top half of his head, containing the Korean text.

방정식
기울기
행렬
룰

운전하는 방법을 배우고 싶은데...

자동차는



어떻게
만들어진 걸까?



A man with dark hair and a blue shirt is driving a car. He has a wide-eyed, shocked expression with his mouth open. His hands are firmly gripping the steering wheel. The background shows the interior of the car and some blurred trees or buildings through the windows, suggesting motion.

엔진 공부 말고, 운전하는 법!

인공지능 프로그래밍 공부하는 법



The place to do data science,
Your home for data science,
Competition and prize money



Competitions

[Documentation](#) [InClass](#)[General](#)[InClass](#)

Sort by

Grouped



All Categories

Search competitions



19 Active Competitions



TWO SIGMA

Two Sigma: Using News to Predict Stock Movements

Use news analytics to predict stock price performance

Featured · Kernels Competition · 5 months to go · 📰 news agencies, time series, finance, money

\$100,000
2,897 teams

LANL Earthquake Prediction

Can you predict upcoming laboratory earthquakes?

Research · 4 months to go · 📰 earth sciences, physics, signal processing

\$50,000
1,059 teams

Elo Merchant Category Recommendation

Help understand customer loyalty

Featured · 16 days to go · 📰 tabular data, banking, regression

\$50,000
3,682 teams

Google Analytics Customer Revenue Prediction

Predict how much GStore customers will spend

Featured · 5 days to go · 📰 regression, tabular data

\$45,000
1,104 teams

 120,303 Datasets

W

Hotness ▾  



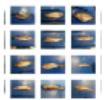
Reddit Vaccine Myths

Gabriel Preda · Updated an hour ago

Usability 10.0 · 1 File (CSV) · 243 kB · 3 Tasks

▲ 1318

Gold ⚙



A Large Scale Fish Dataset

Oğuzhan Ulucan · Updated 7 months ago

Usability 9.4 · 18436 Files (other) · 3 GB · 2 Tasks

▲ 732

Gold ⚙



MusicNet Dataset

Sparsh Gupta · Updated 9 months ago

Usability 10.0 · 992 Files (other, CSV) · 23 GB

▲ 330

Gold ⚙



Wikibooks Dataset

Dhruvil Dave · Updated a month ago

Usability 10.0 · 1 File (SQLITE) · 2 GB · 1 Task

▲ 287

Gold ⚙



Careerbuilder Job Listing 2020

PromptCloud · Updated 9 months ago

Usability 10.0 · 1 File (other) · 44 MB · 1 Task

▲ 102

Bronze ⚙



Famous Iconic Women

Fatima-Ezzahra · Updated 9 months ago

Usability 7.5 · 3146 Files (other) · 879 MB · 1 Task

▲ 145

Silver ⚙

Kernels

[Documentation](#)[New Kernel](#)

Earn bragging rights (plus a \$2000 award!) when you share your data science project on Kaggle and are chosen as our Kernel Author of the Month!

[Public](#)[Your Work](#)[Favorites](#)

Sort by

Hotness

Categories

Outputs

Languages

Types

Tags

Search kernels



2441

[Titanic Data Science Solutions](#)

7h ago in titanic • tutorial, feature engineering, model comparison



Py

707

71

[target - true meaning revealed!](#)

8h ago in elo-merchant-category-recommendation



Py

11

18

[How to not overfit?](#)

4h ago in dont-overfit-ii • 0.848 • beginner, data visualization, classification, starter code, model ...



Py

4

109

[LANL Earthquake EDA and Prediction](#)

12h ago in LANL-Earthquake-Prediction • 1.456



Py

44

16

[TMDB EDA](#)

1h ago with multiple data sources • 1.94744



Py

0

github.com/yungbyun

/female-male-classification-original