

기계학습개론

02 Python 기본 문법, Numpy,
Pandas 개요

컴퓨터공학과 이상금

N4-416

sangkeum@hanbat.ac.kr



1. 공지사항

한밭대학교 SW중심대학사업단
유튜브 구독



1. 공지사항



The poster features a blue background with white and yellow text. At the top, it says '데이터안심구역 활용 2023 공동경진대회'. Below this, it lists prizes: '과기정통부장관상 등 상훈 총 15점과 상금 총 2,200만원! 수상자 채용 특전까지!'. It then states that anyone interested in data security in South Korea can participate, and a '2023 데이터안심구역 활용 공동 경진대회 설명회' will be held. The event details are: Date: September 12 (Wed) 14:00-16:00, Location: Chungnam National University Library Lecture Hall (Changwon Campus). The application period is from September 1 to September 11. At the bottom, there is a large yellow button that says '설명회 참가신청'. Decorative elements include a 3D bar chart with data labels like '40%', '57%', and '100%' on the right side, and stylized geometric shapes in the corners.

데이터안심구역 활용 2023
공동경진대회

과기정통부장관상 등 상훈 총 15점과
상금 총 **2,200만원!**
수상자 채용 특전까지!

데이터 활용에 관심있는 대한민국 국민이라면 누구나 참여 가능한
2023 데이터안심구역 활용 공동 경진대회 설명회를 개최합니다!

- 일시: 9월 12일(화) 14:00 - 16:00
- 장소: 충남대학교 도서관 강당 (대전시 유성구 소재)

설명회 참가 신청기간: 9.1.(금) ~ 9.11.(월)

설명회 참가신청

1. 강의계획

- 목표

데이터분석(NumPy, Pandas)과 기계학습(Scikit-Learn) 학습

- 참고문헌

(메인) 혼자 공부하는 머신러닝+딥러닝
데이터 과학을 위한 파이썬 머신러닝

- 평가방법

출석: 15%

과제 및 퀴즈: 15%

중간고사: 30%

기말고사: 40%

- 담당TA

함규식 석사과정

이메일: 30231218@edu.hanbat.ac.kr

한보연 석사과정

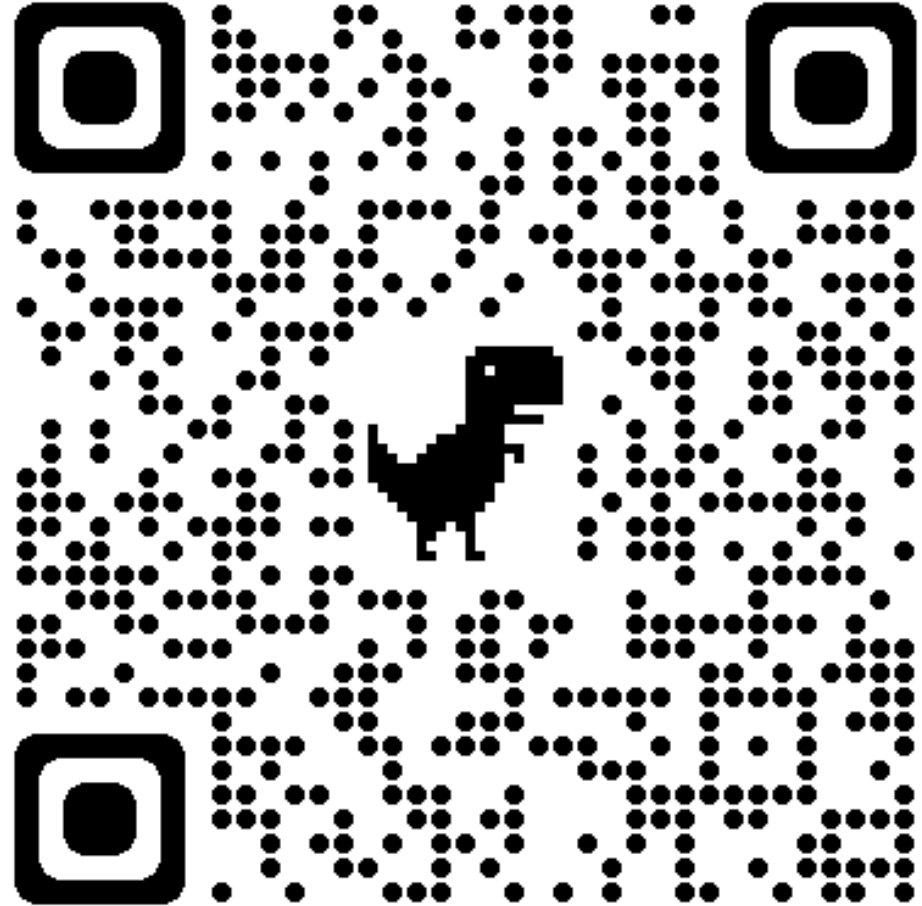
이메일: gksqhdus2@naver.com

*추후 점수 및 수업내용 변경 가능

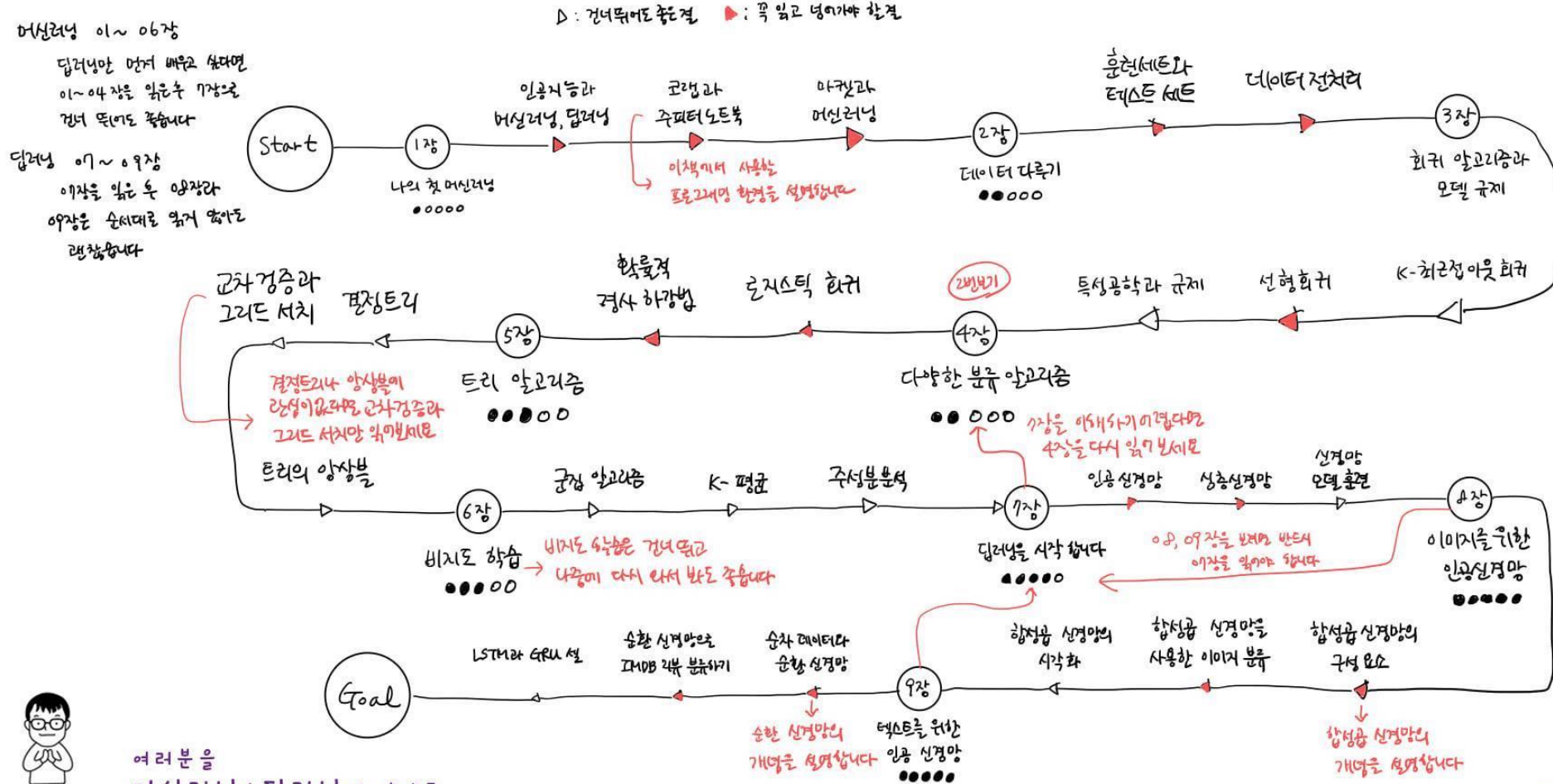
주차	강의계획
1주차	머신러닝 개요 및 기본문법
2주차	Python 기본 문법, Numpy, Pandas 개요
3주차	머신러닝 프로젝트 처음부터 끝까지
4주차	머신러닝 프로젝트 처음부터 끝까지
5주차	분류
6주차	분류(2)
7주차	중간고사
8주차	모델훈련
9주차	모델훈련(2)
10주차	서포트벡터 머신
11주차	머신러닝 분야 실무 적용 사례
12주차	결정트리
13주차	랜덤포레스트
14주차	랜덤포레스트(2)
15주차	기말고사

1. 강의계획

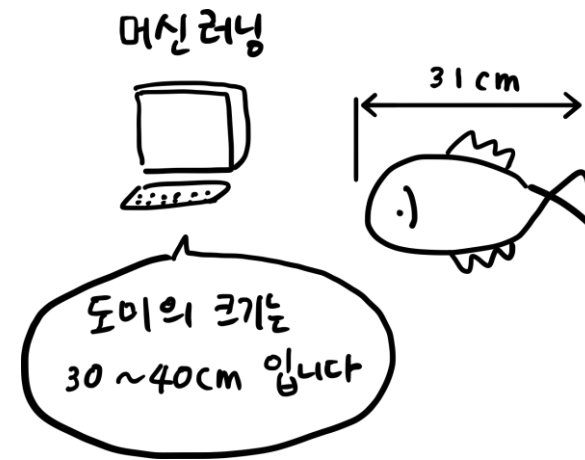
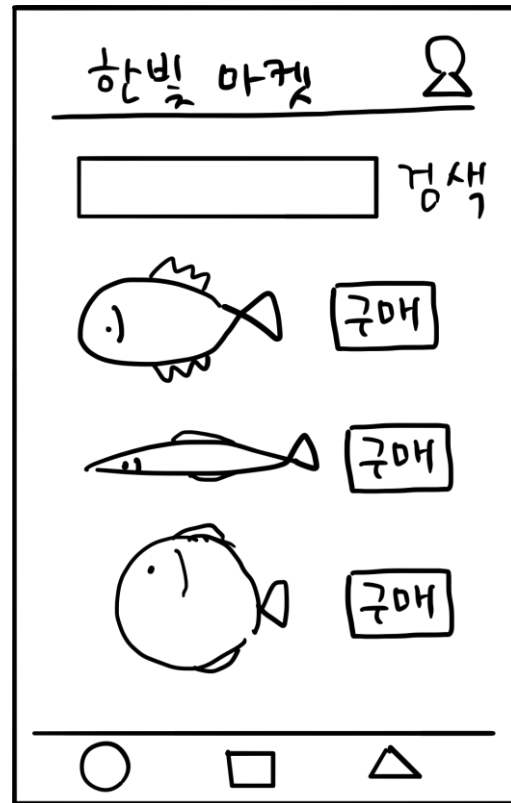
- 설문조사
수업에서 바라는 점 or 저에게 한마디
익명으로 의견 작성 (10분), 휴대폰으로
QR코드 인식
- 피드백 진행



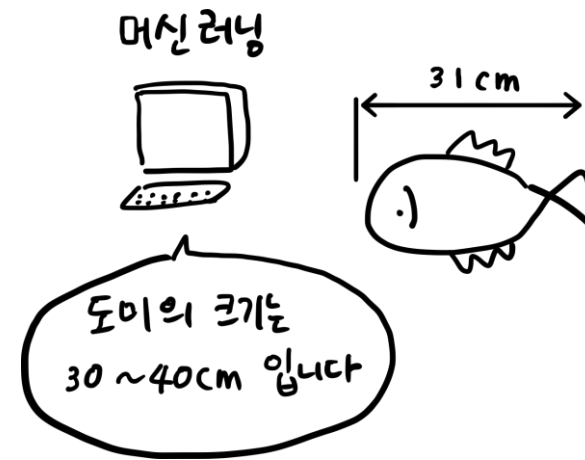
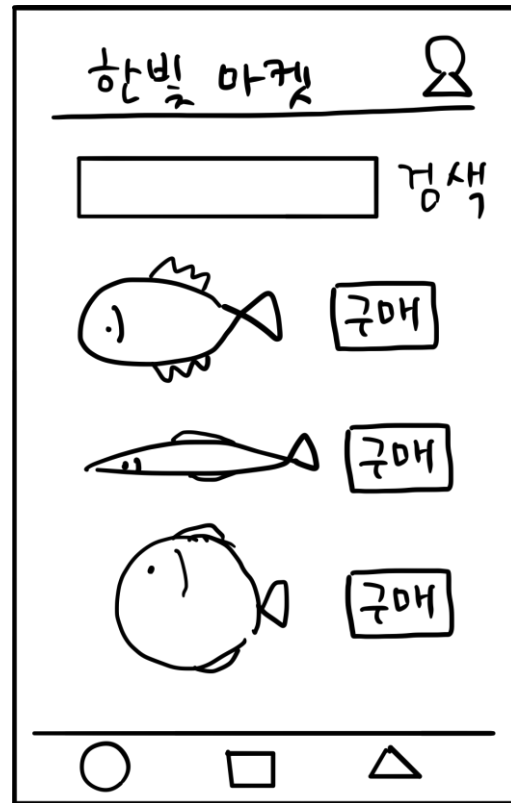
1. 강의계획



2. 첫 번째 머신러닝 프로그램

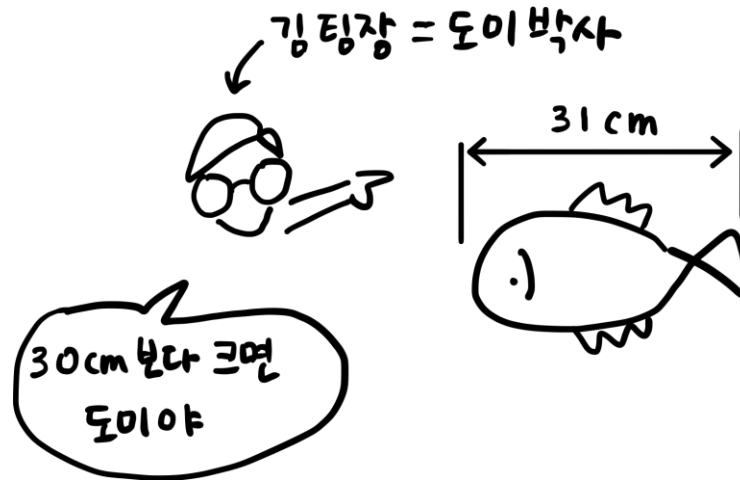


2. 첫 번째 머신러닝 프로그램



2. 첫 번째 머신러닝 프로그램

전통적인 프로그램



```
if fish_length >= 30:  
    print("도미")
```

2. 첫 번째 머신러닝 프로그램

도미 vs 빙어

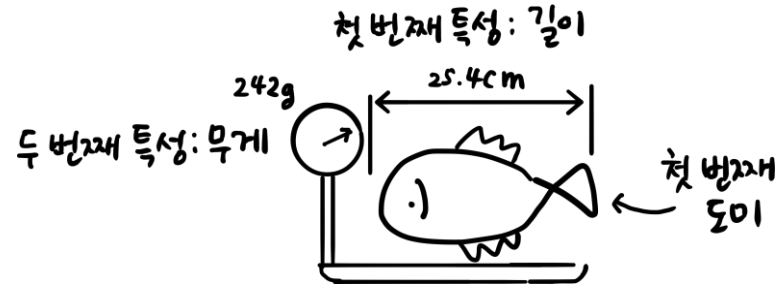
2개의 클래스(class)

분류(classification)

이진 분류(binary classification)

2. 첫 번째 머신러닝 프로그램

도미 데이터



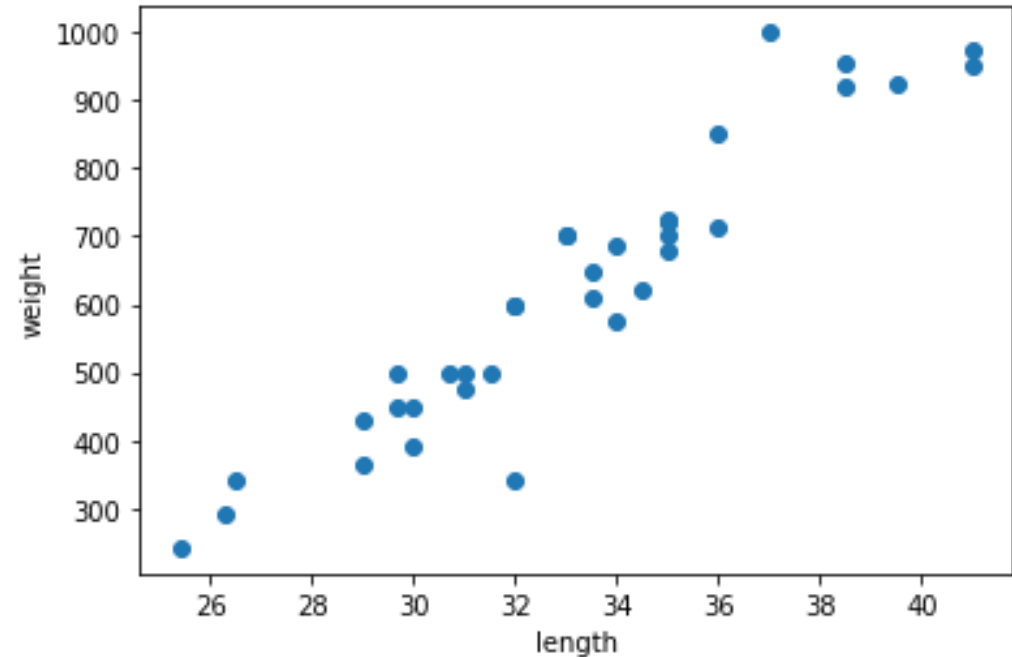
```
bream_length = [25.4, 26.3, 26.5, 29.0, 29.0, 29.7, 29.7, 30.0, 30.0, 30.7,
                31.0, 31.0, 31.5, 32.0, 32.0, 32.0, 33.0, 33.0, 33.5, 33.5,
                34.0, 34.0, 34.5, 35.0, 35.0, 35.0, 35.0, 36.0, 36.0, 37.0,
                38.5, 38.5, 39.5, 41.0, 41.0]
bream_weight = [242.0, 290.0, 340.0, 363.0, 430.0, 450.0, 500.0, 390.0,
                450.0, 500.0, 475.0, 500.0, 500.0, 340.0, 600.0, 600.0,
                700.0, 700.0, 610.0, 650.0, 575.0, 685.0, 620.0, 680.0,
                700.0, 725.0, 720.0, 714.0, 850.0, 1000.0, 920.0, 955.0,
                925.0, 975.0, 950.0]
```

2. 첫 번째 머신러닝 프로그램

산점도(scatter plot)

```
import matplotlib.pyplot as plt

plt.scatter(bream_length, bream_weight)
plt.xlabel('length')
plt.ylabel('weight')
plt.show()
```

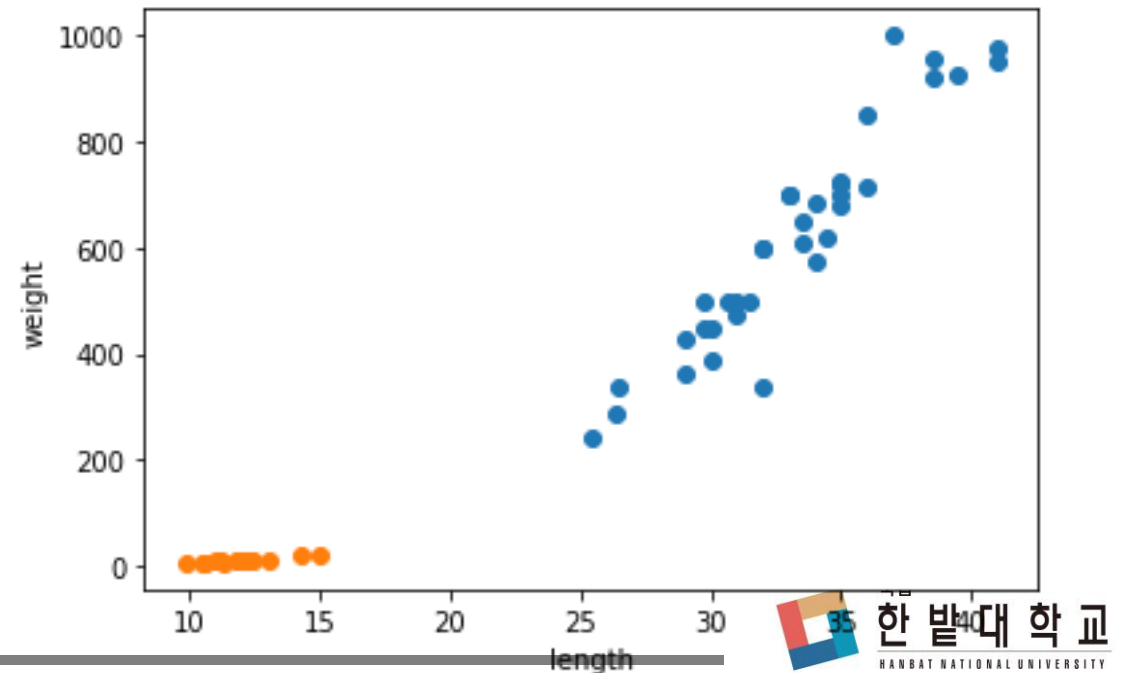


2. 첫 번째 머신러닝 프로그램

빙어 데이터

```
smelt_length = [9.8, 10.5, 10.6, 11.0, 11.2, 11.3, 11.8, 11.8, 12.0, 12.2,  
               12.4, 13.0, 14.3, 15.0]  
smelt_weight = [6.7, 7.5, 7.0, 9.7, 9.8, 8.7, 10.0, 9.9, 9.8, 12.2, 13.4,  
               12.2, 19.7, 19.9]
```

```
plt.scatter(bream_length, bream_weight)  
plt.scatter(smelt_length, smelt_weight)  
plt.xlabel('length')  
plt.ylabel('weight')  
plt.show()
```



2. 첫 번째 머신러닝 프로그램

도미와 빙어 합치기

```
length = bream_length+smelt_length  
weight = bream_weight+smelt_weight
```

도미 35개의 길이 빙어 14개의 길이

length = [25.4, 26.3, ... , 41.0, 9.8, ... , 15.0]

도미 35개의 무게 빙어 14개의 무게

weight = [242.0, 290.0, ... , 950.0, 6.7, ... , 19.9]



사이킷런이 기대하는 데이터 형태

길이 무게

49개의 생선 {

[25.4,	242.0],
[26.3,	290.0],
.	.
.	.
.	.
[15.0,	19.9]]

2. 첫 번째 머신러닝 프로그램

리스트 내포

```
fish_data = [[l, w] for l, w in zip(length, weight)]
```

```
[[25.4, 242.0], [26.3, 290.0], [26.5, 340.0], [29.0, 363.0], [29.0, 430.0],  
[29.7, 450.0], [29.7, 500.0], [30.0, 390.0], [30.0, 450.0], [30.7, 500.0],  
[31.0, 475.0], [31.0, 500.0], [31.5, 500.0], [32.0, 340.0], [32.0, 600.0],  
[32.0, 600.0], [33.0, 700.0], [33.0, 700.0], [33.5, 610.0], [33.5, 650.0],  
[34.0, 575.0], [34.0, 685.0], [34.5, 620.0], [35.0, 680.0], [35.0, 700.0],  
[35.0, 725.0], [35.0, 720.0], [36.0, 714.0], [36.0, 850.0], [37.0, 1000.0],  
[38.5, 920.0], [38.5, 955.0], [39.5, 925.0], [41.0, 975.0], [41.0, 950.0],  
[9.8, 6.7], [10.5, 7.5], [10.6, 7.0], [11.0, 9.7], [11.2, 9.8], [11.3, 8.7],  
[11.8, 10.0], [11.8, 9.9], [12.0, 9.8], [12.2, 12.2], [12.4, 13.4],  
[13.0, 12.2], [14.3, 19.7], [15.0, 19.9]]
```

2. 첫 번째 머신러닝 프로그램

정답준비

```
fish_target = [1]*35 + [0]*14
```

[1,
1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

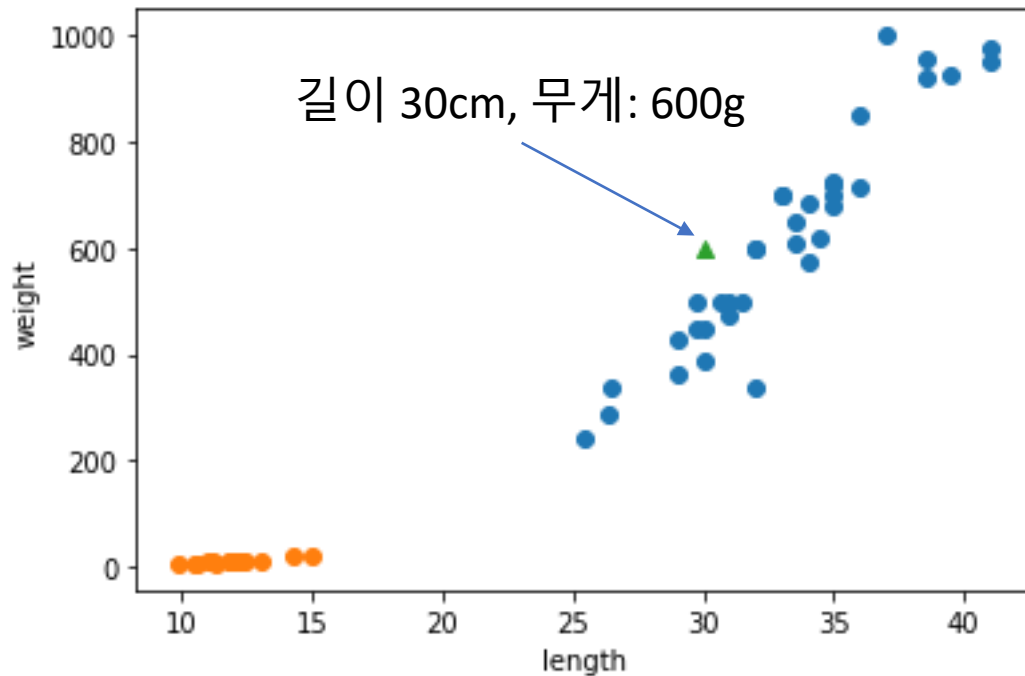
2. 첫 번째 머신러닝 프로그램

k-최근접 이웃

```
from sklearn.neighbors import KNeighborsClassifier  
  
kn = KNeighborsClassifier()  
  
kn.fit(fish_data, fish_target)  
  
kn.score(fish_data, fish_target)  
1.0
```

2. 첫 번째 머신러닝 프로그램

새로운 생선 예측



```
kn.predict([[30, 600]])
```

```
array([1])
```

2. 첫 번째 머신러닝 프로그램

무조건 도미

```
kn49 = KNeighborsClassifier(n_neighbors=49)
```

```
kn49.fit(fish_data, fish_target)  
kn49.score(fish_data, fish_target)
```

```
0.7142857142857143
```

```
print(35/49)
```

```
0.7142857142857143
```

Q&A
감사합니다