k-Nearest Neighbor

Code Examples

부산대학교 정보·의생명공학대학
정보컴퓨터공학부
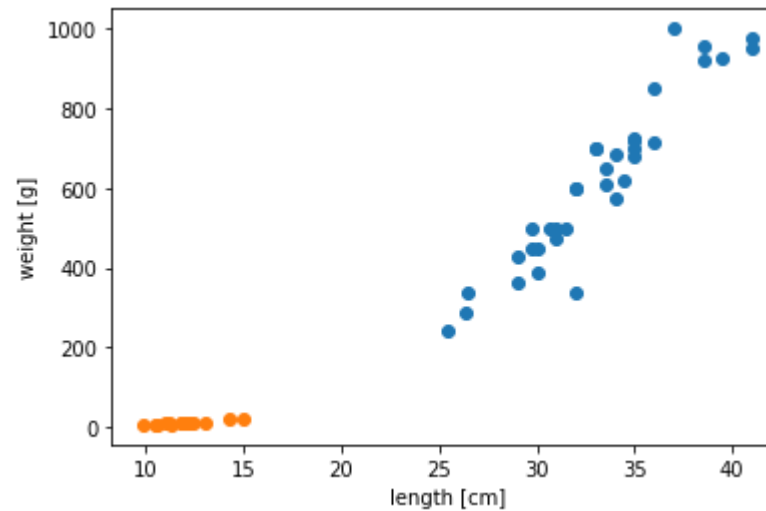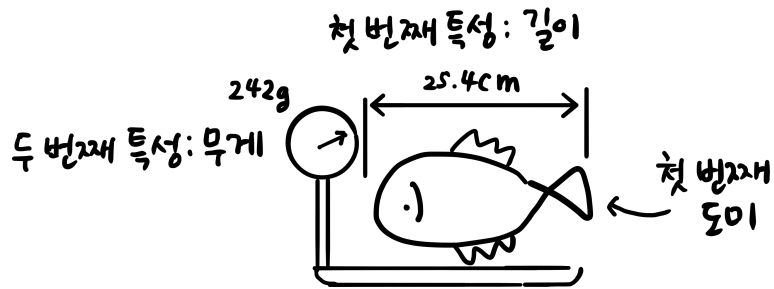
❖ Bream(도미) or Smelt(빙어)

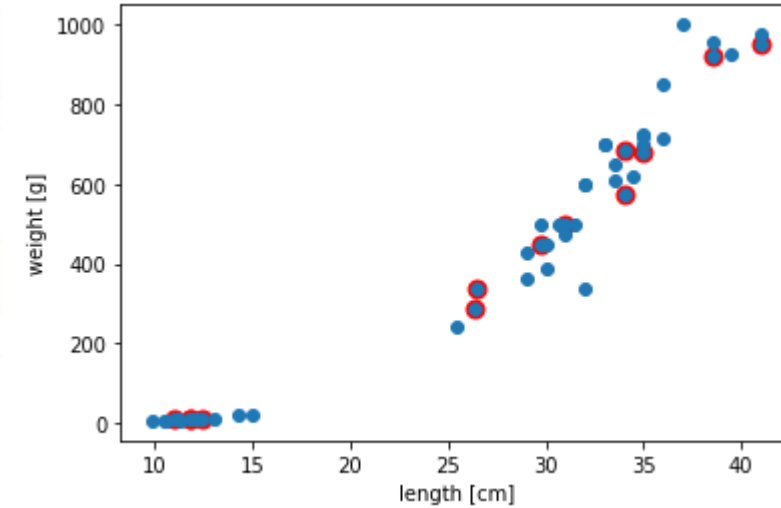 ▪ Binary Classification

❖ Features: Length and Weight
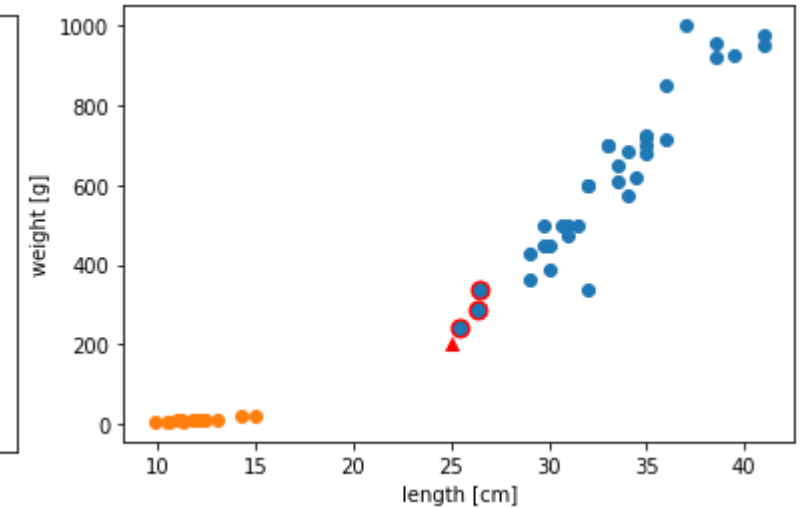

```python
from sklearn.neighbors import KNeighborsClassifier

kn = KNeighborsClassifier(n_neighbors=3, metric='euclidean')
kn.fit(fish_data, fish_target)
```
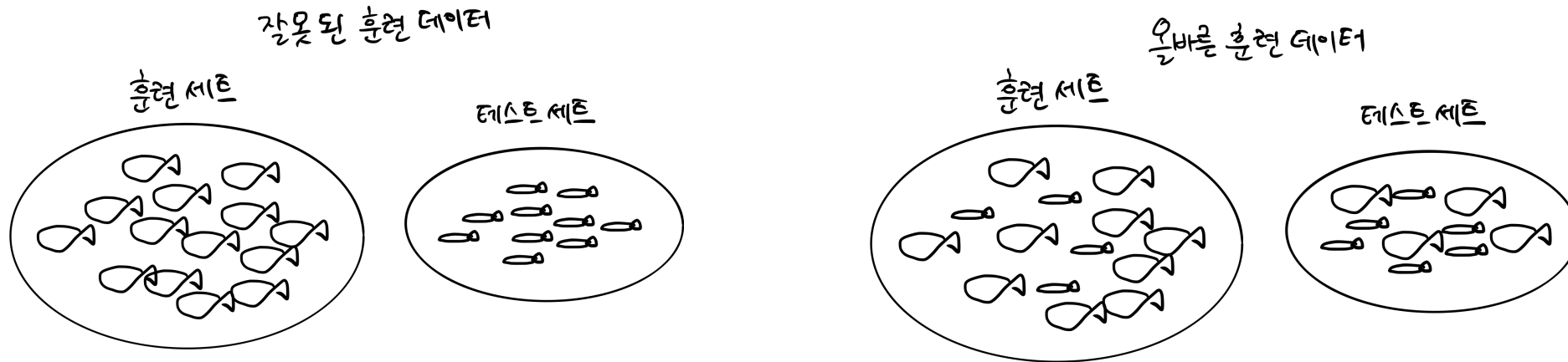
```python
newFish = [25, 200]
kn.predict([newFish])
```

kNN Example – Data Splitting

❖ Training set, test set

A | Training | Test |

Single Dataset

B | Training | Validation | Test |

Single Dataset

❖ Sampling bias

잘못된 훈련 데이터

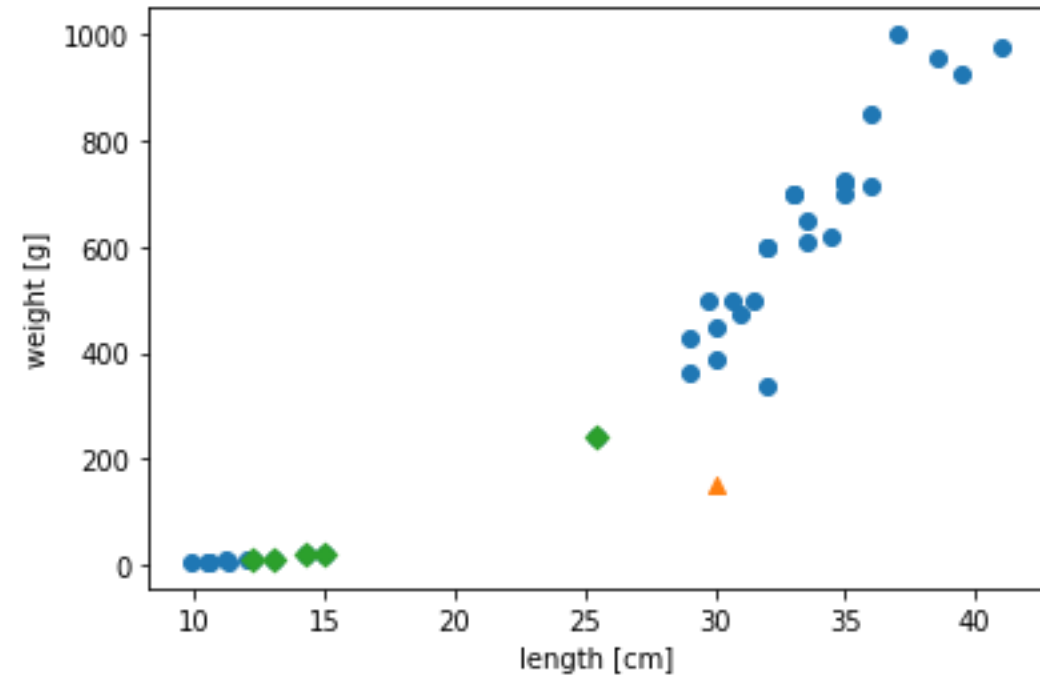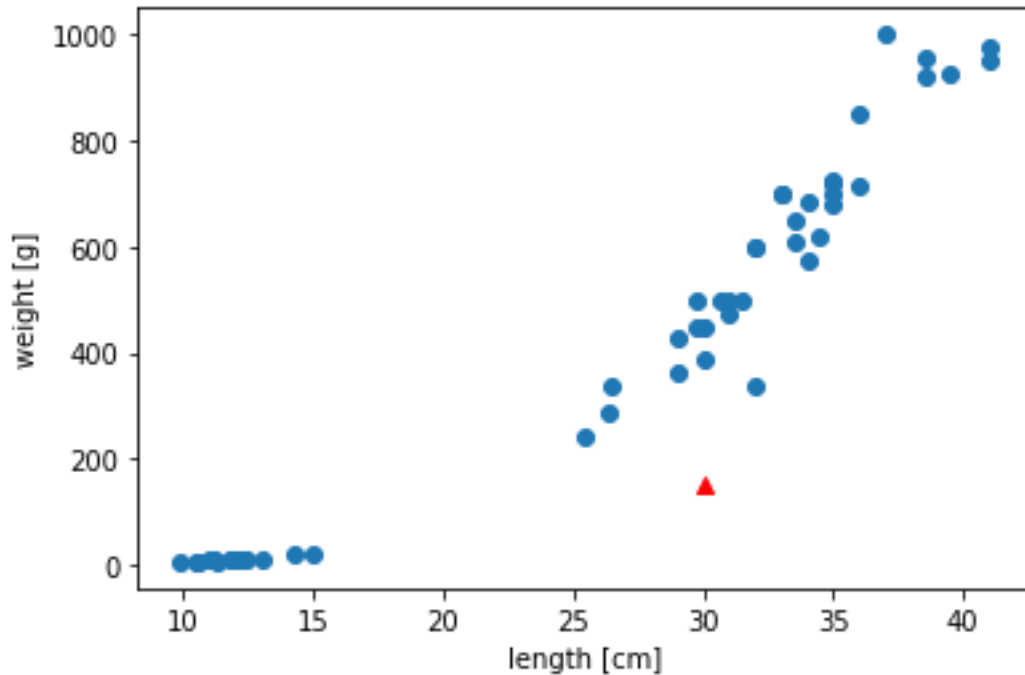훈련 세트　　　테스트 세트

올바른 훈련 데이터

훈련 세트　　　테스트 세트

```
train_input, test_input, train_target, test_target = train_test_split(
    fish_data, fish_target, stratify=fish_target, random_state = 42)
```

kNN Example – Feature scaling

❖ Unknown fish: 30 cm and 150 g

❖ Bream or Smelt?

kNN Example – solutions for scale issue

❖ Data preprocessing:

  ▪ Standardization

```python
mean = np.mean(train_input, axis=0)
std = np.std(train_input, axis=0)
train_scaled = (train_input - mean) / std
```

```python
kn.fit(train_scaled, train_target)
```

```python
newFish_scaled = (newFish - mean) / std
print(kn.predict([newFish_scaled]))
```

❖ Distance metric

```python
kn = KNeighborsClassifier(algorithm='brute', metric='mahalanobis'
, metric_params={'VI': np.linalg.inv(np.cov(train_input.T)).T})
kn.fit(train_input, train_target)
```