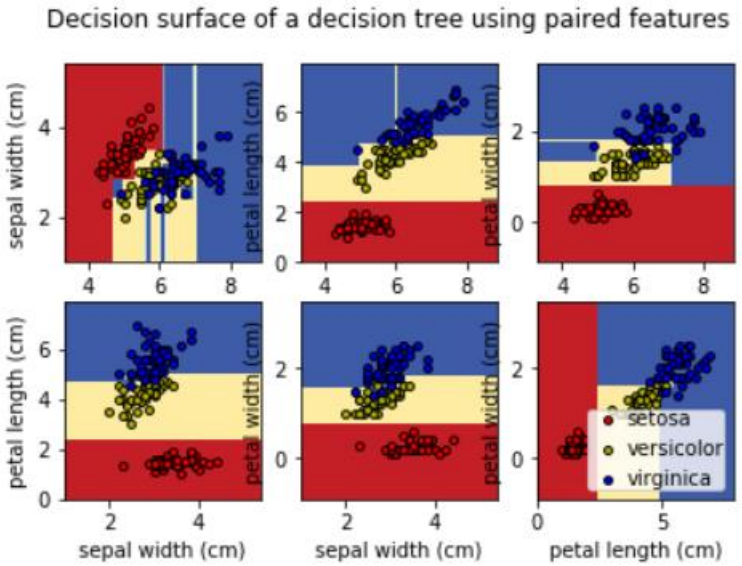
동국대학교
dongguk university

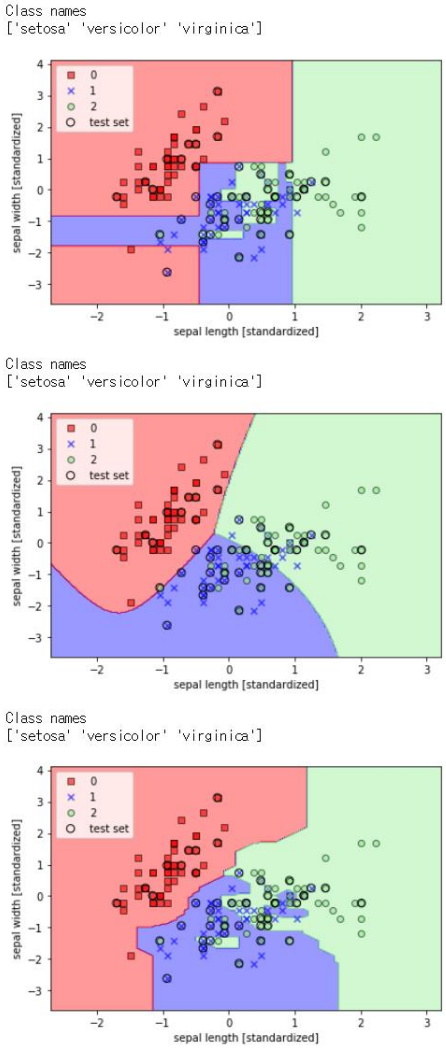
Machin Learning

Practice5

실습 1. Decision Region of Decision Tree

- Sklearn
  - Make a decision Region of Decision Tree





실습 2. Custom Decision Tree Implementation

- Make a Entropy Function and Information gain Function

Exercise 1 make a entropy function

```
def entropy(rows):
    """Calculate the Entropy for a list of rows."""
    #make a entropy


#######
# Demo:
# Let's look at some example to understand how Entropy works.
no_mixing = [['Apple'],
             ['Apple'],
             ['Apple'],
             ['Apple']]
print("1. no mixing: ", entropy(no_mixing))

some_mixing = [['Apple'],
               ['Apple'],
               ['Apple'],
               ['Orange']]
print("2. some mixing: ", entropy(some_mixing))

half_and_half = [['Apple'],
                 ['Apple'],
                 ['Orange'],
                 ['Orange']]
print("3. half mixing: ", entropy(half_and_half))


lots_of_mixing = [['Apple'],
                  ['Orange'],
                  ['Grape'],
                  ['Grapefruit'],
                  ['Blueberry']]
print("4. lots of mixing: ", entropy(lots_of_mixing))
#######
```

```
1. no mixing:  0.0
2. some mixing:  0.8112781244591328
3. half mixing:  1.0
4. lots of mixing:  2.321928094887362
```

Exercise 2 make a info_gain function using entropy

```
def info_gain(left, right, current_entropy):
    """Information Gain."""
    #make a information gain


#######
# Demo:
# Calculate the uncertainy of our training data.
current_entropy = entropy(training_data)
print("1. current entropy: ", current_entropy)


# How much information do we gain by partioning on 'Green'?
true_rows, false_rows = partition(training_data, Question(0, 'Green'))
print("2. informatino gain: ", info_gain(true_rows, false_rows, current_entropy))


# What about if we partioned on 'Red' instead?
true_rows, false_rows = partition(training_data, Question(0,'Red'))
print("3. informatino gain: ", info_gain(true_rows, false_rows, current_entropy))
```

```
1. current entropy:  1.5219280948873621
2. informatino gain:  0.32192809488736196
3. informatino gain:  0.9709505944546685
```

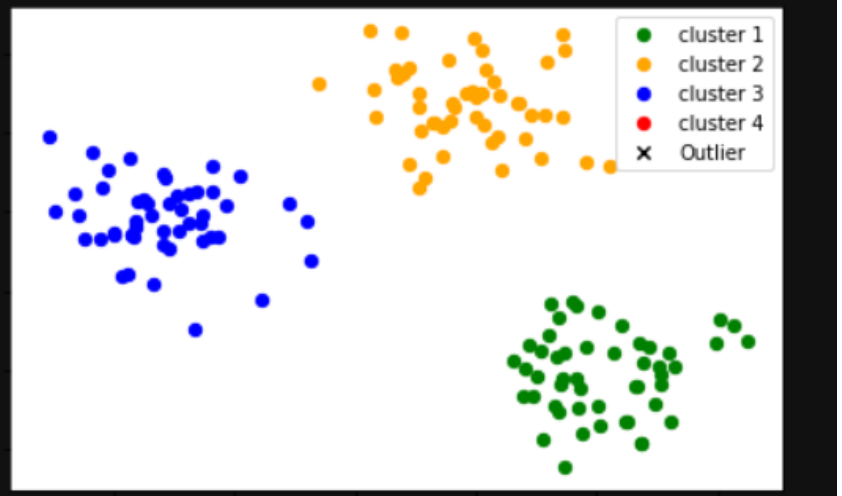
실습 3. Clustering

- Outlier Detection using DBSCAN

Exercise 1

No Outlier

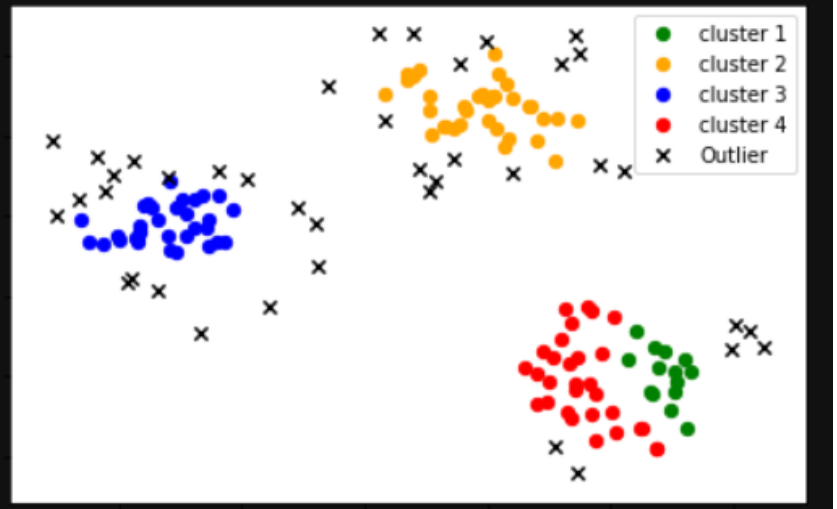
```
# Make a Plot
```



Exercise 2

Bad Clustering

```
# Make a Plot
```



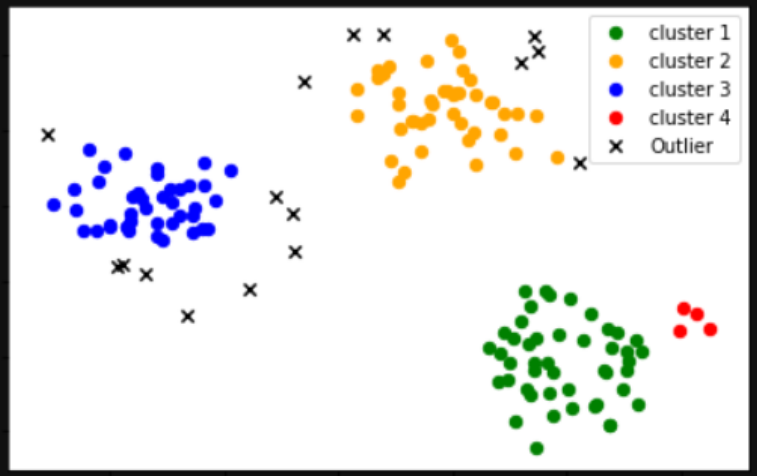
실습 3. Clustering

- Outlier Detection using DBSCAN

Exercise 3

There are few outlier but they have 4 clusters

```
# Make a Plot
```



Exercise 4

There are few outlier

```
# Make a Plot
```