

13주차 과제

Classifier



담당교수님: 윤장혁 교수님

과목명: Data Analytics

이름: 박민성

학번: 201611145

전공: 산업공학과

제출일: 2020.06.09

Object

- 주어진 데이터(data_week13.csv)를 이용하여 classifier 학습
 - 주어진 데이터는 4개의 input variable과 1개의 output variable로 구성됨
 - Input variable은 꽃받침 길이 및 너비, 꽃잎의 길이 및 너비임
 - output variable은 꽃의 종류로 setosa / versicolor / virginica 3종류로 구분됨
- Decision Tree를 활용
 - 최종 Tree는 test accuracy 기준 55% 이상을 만족시켜야 함
 - Max depth는 5 이하로 설정 (성능만 충족한다면 반드시 5일 필요는 없음)
 - 최종 Tree의 Diagram 혹은 영역이 구분되어 있는 그래프 중 하나의 그림을 반드시 포함할 것

Classifier Learning, Decision Tree

```
C: > Users > minisong > Desktop > 13주차.py > ...
1  import pandas as pd
2  import numpy as np
3  from sklearn.metrics import accuracy_score
4  from sklearn.model_selection import train_test_split
5  from sklearn.preprocessing import StandardScaler
6  import os
7  from sklearn import tree
8  import graphviz
9  from sklearn.tree import export_graphviz
10 os.environ["PATH"] += os.pathsep + 'C:/Program Files (x86)/Graphviz2.38/bin/'
11
12 df=pd.read_csv('./data_week13.csv')
13 Y=df['Species']
14 X=df[['SepalLength', 'SepalWidth', 'PetalLength', 'PetalWidth']]
15
16 clf = tree.DecisionTreeClassifier(criterion='entropy',max_depth=5)
17 clf = clf.fit(X, Y)
18 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.3)
19 clf.fit(X_train,Y_train)
20 print(round(clf.score(X_test,Y_test),2)*100,"%")
21
22 class_names = clf.classes_
23
24 export_graphviz(clf,out_file='result.dot',
25                 class_names= class_names,
26                 feature_names=['SepalLength', 'SepalWidth', 'PetalLength', 'PetalWidth'],
27                 impurity=False, filled=True)
28
29 with open('result.dot') as file_reader:
30     dot_graph = file_reader.read()
31     dot = graphviz.Source(dot_graph)
32     dot.render(filename='C:\\Users\\minisong\\Desktop\\결과.png')
```

<분석을 진행한 파이썬 코드>

data_week13.csv에서 data load 후 의사결정나무 모델을 학습시켜서 테스트를 진행했다.

분석은 파이썬으로 진행하였다.

Input Variable은 SepalLength, SepalWidth, PetalLength, PetalWidth 4가지이고

Output Variable은 Species이다.

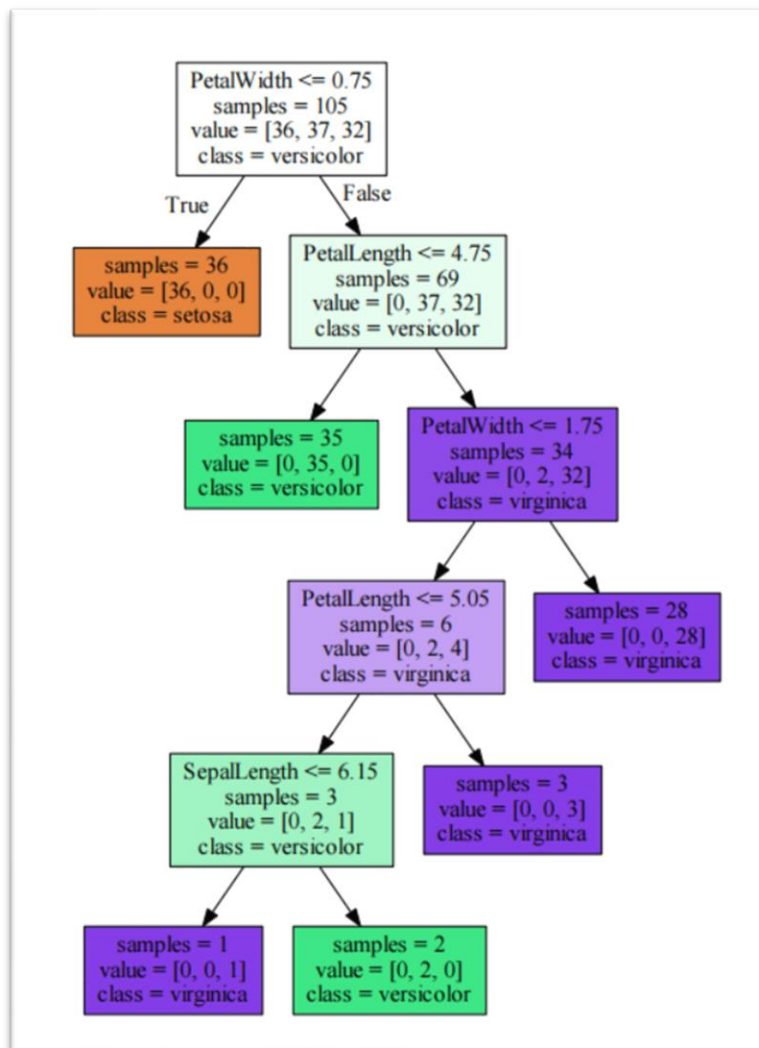
93.0 %

Test size: 0.3, Max Depth: 5으로 설정했다.

Test 정확도: 약 93%으로 나왔고 55%를 넘었기에 기준을 만족했다.

파이썬 모듈을 이용하여 의사결정나무를 시각화하였고, 결과.pdf로 그림으로 저장하여 나타냈다.

Tree의 Diagram Result



<결과.pdf>