

- 교과서 244페이지를 참고하여 scikit learn 패키지를 설치합니다.
- 아래 첨부된 코드를 작성하고, 실행합니다. 이 코드는 교과서 264-265 페이지에 있는 코드와 동일한 코드입니다.
  - 어려우신 부분들은 교과서 253쪽부터 천천히 읽어보시면서 이해하시면 됩니다.
  - 그런데 코드를 모두 이해하지 못하셔도 괜찮습니다. 코드 분석하는 과제는 아니니까요.

```
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn import tree
from sklearn.metrics import accuracy_score

def plot_accuracy(prediction, truth):

    acc = accuracy_score(prediction, truth)
    print (acc)

iris = datasets.load_iris()

X = iris.data
y = iris.target

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.4)

clf = tree.DecisionTreeClassifier()
clf.fit(X_train, y_train)

predictions = clf.predict(X_test)

print('예측:\n', predictions)
print('정답(y_test):\n', y_test)

plot_accuracy(predictions, y_test)
```

- 과제입니다.

## 1. 첨부된 코드를 실행하고, 아래 내용에 맞춰 보고서를 작성합니다.

- 아래 코드를 보면 `test_size=.4`로 되어 있습니다. 이 값을 각각 0.7, 0.5, 0.3으로 바꾸고 실행합니다.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.4)
```

- [0.7, 0.5, 0.3]에 대해 각각 prediction accuracy가 출력될 것입니다 (다시 말해, 값을 넣고 한번씩 실행해서 결과를 모으라는 말씀입니다).
- X를 [0.7, 0.5, 0.3], Y를 [accuracy 1, accuracy 2, accuracy3] – 다시 말해 [0.7일때 accuracy, 0.5일때 accuracy, 0.3일때 accuracy]로 구성하고 bar plot을 그립니다.
- Accuracy는 왜 다르게 나올까요? 여러분들께서 생각하시는 이유를 자유롭게 기술해 봅시다.
- 그래프와 함께 상기 이유를 기술하고 제출합니다. 보고서의 형식대로 작성하여 제출합니다.
- Bar plot을 그리는 소스가 포함된 전체 소스코드 1식, 보고서 1식, 총 2개 파일을 제출해 주세요.