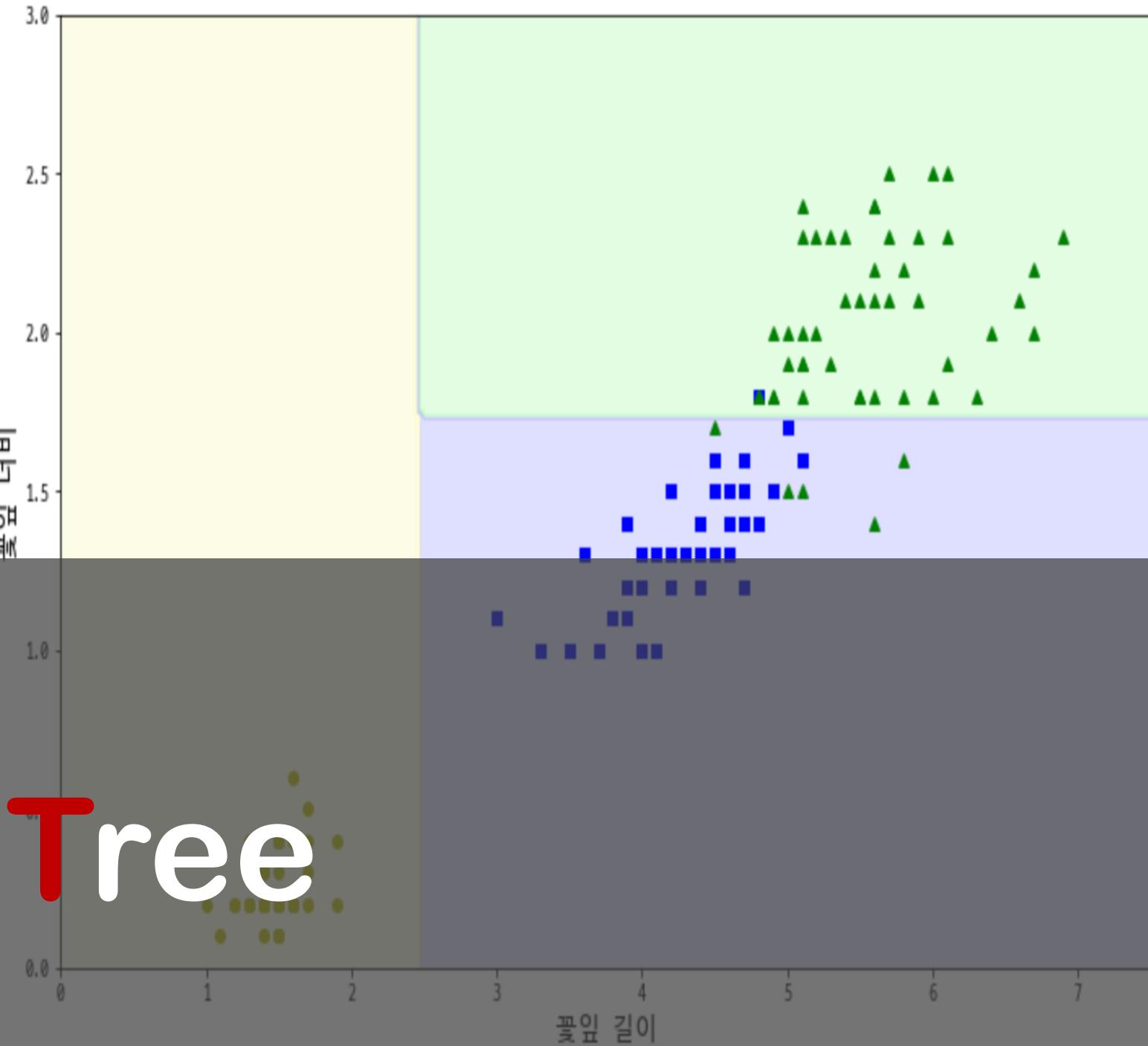
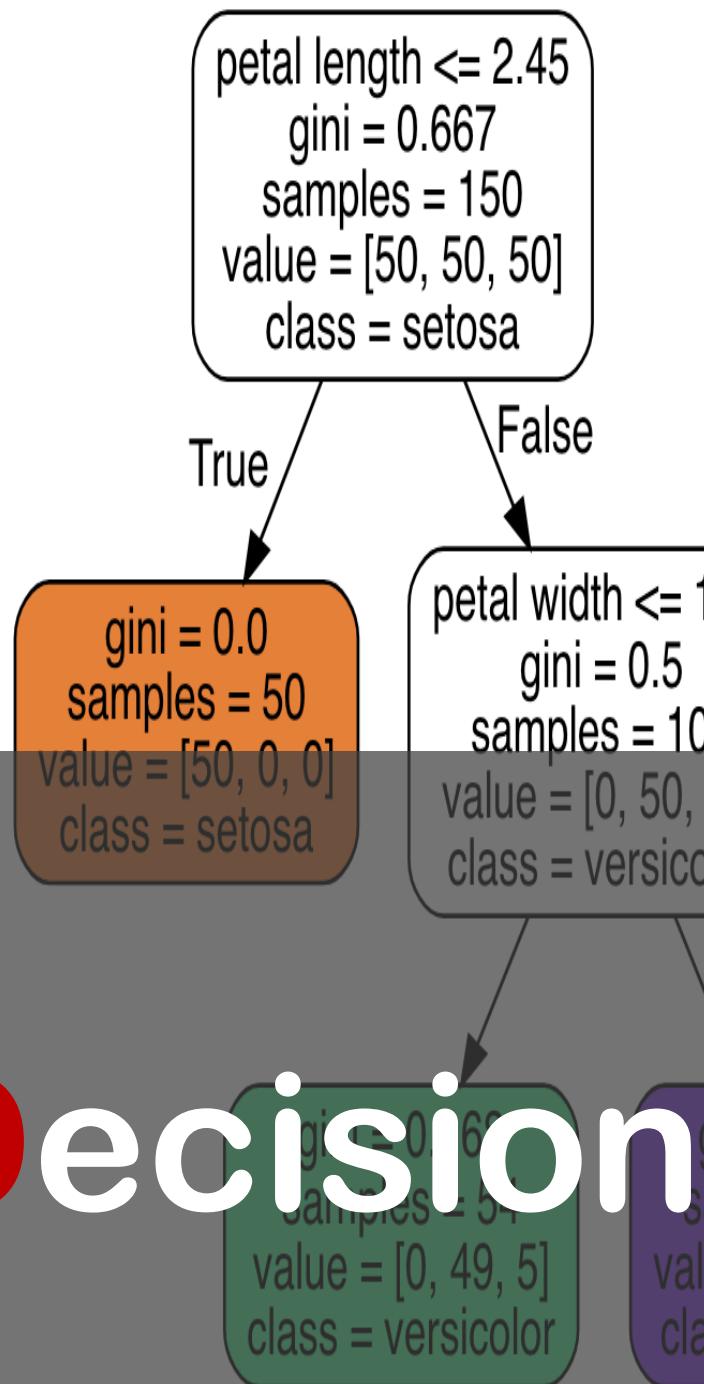


Decision Tree



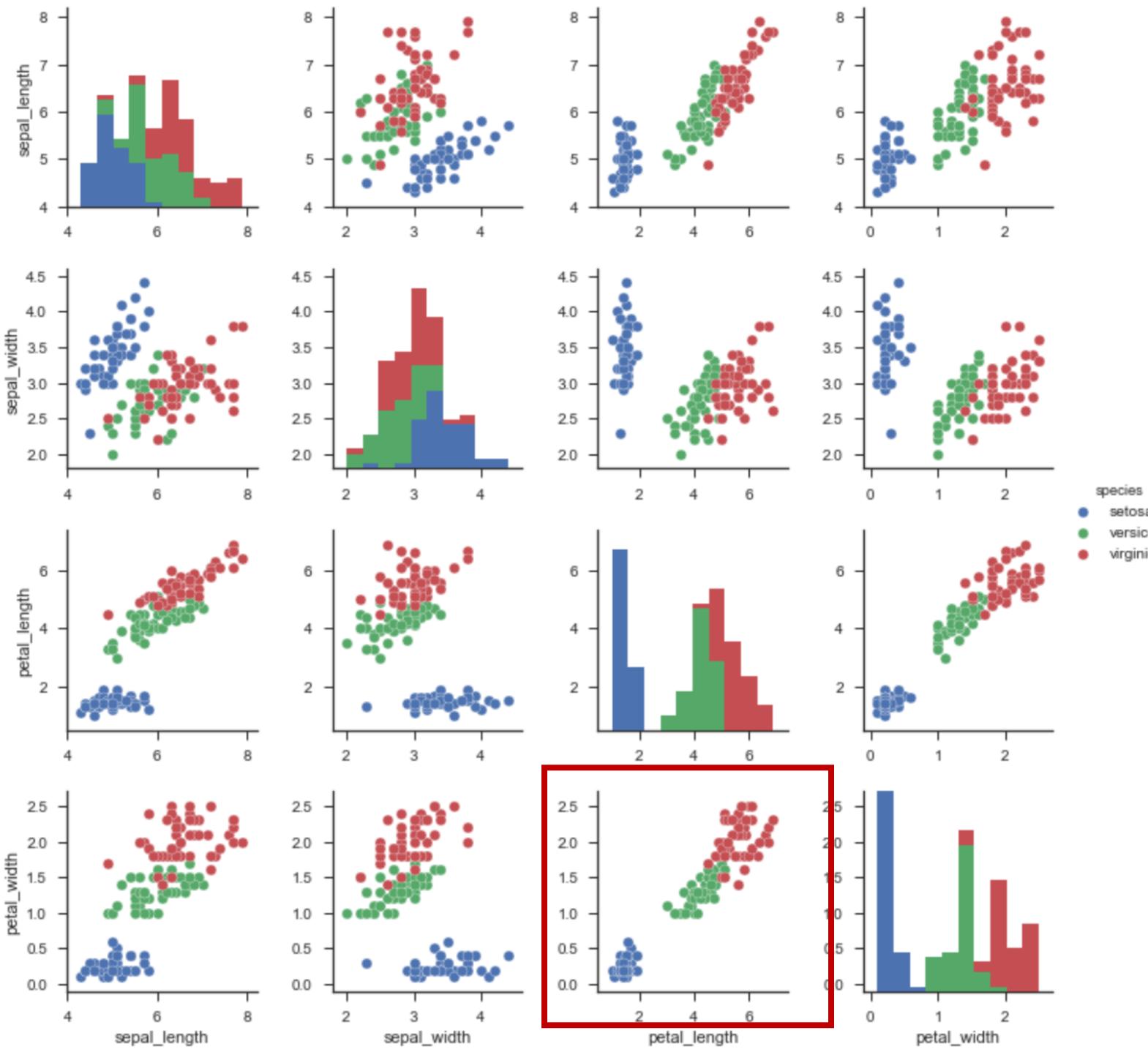
-
- 혹시.. 3주차때 했던 **Seaborn** 투토리얼에서
 - 붓꽃 **Iris** 그림 기억 나나요????

Seaborn – Data Visualization



Seaborn – Data Visualization

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa
8	4.4	2.9	1.4	0.2	setosa
9	4.9	3.1	1.5	0.1	setosa



In [28]:

```
from sklearn.datasets import load_iris
from sklearn.tree import DecisionTreeClassifier

iris = load_iris()
X = iris.data[:, 2:] # petal length and width
y = iris.target
```

In [30]:

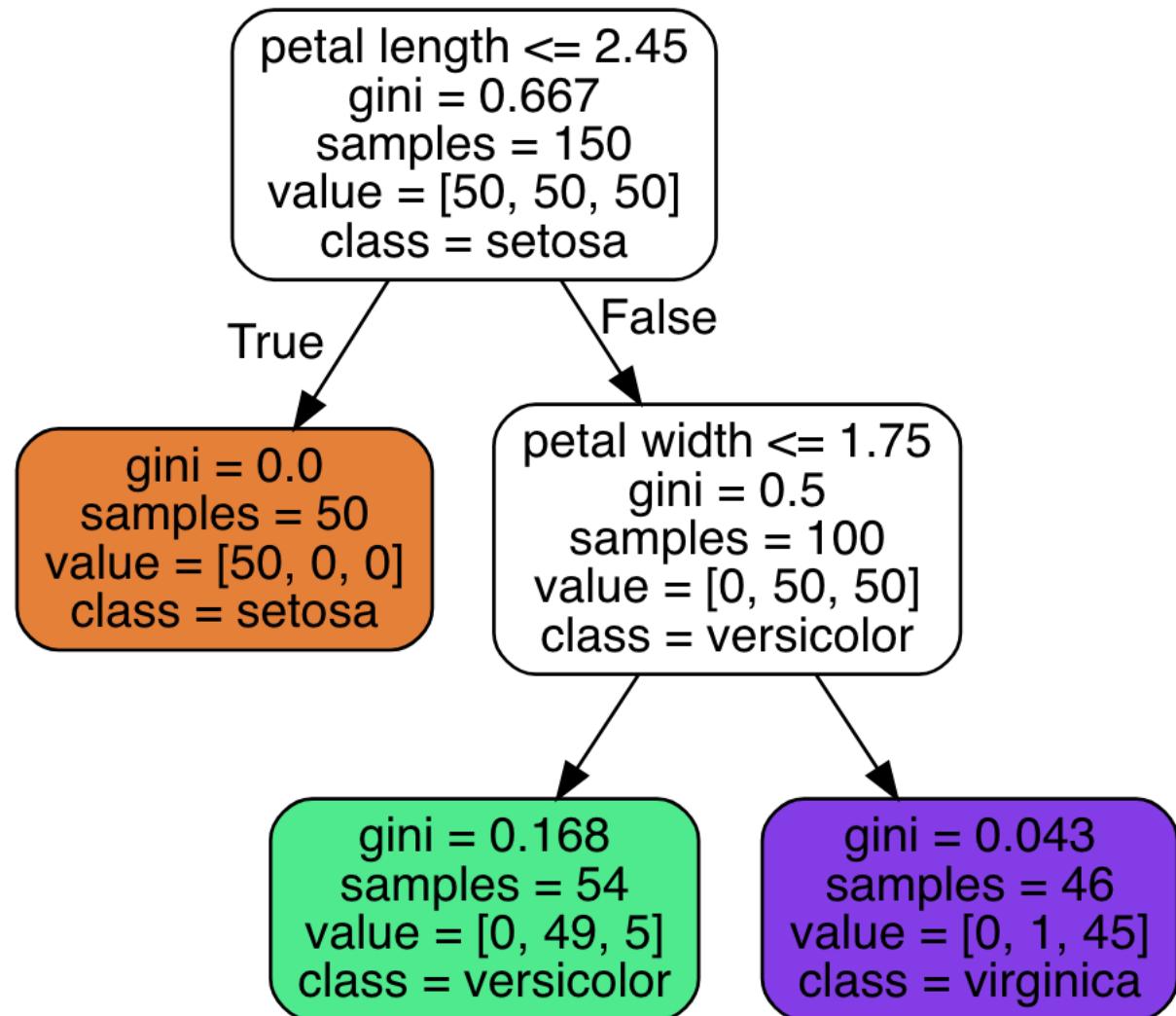
x

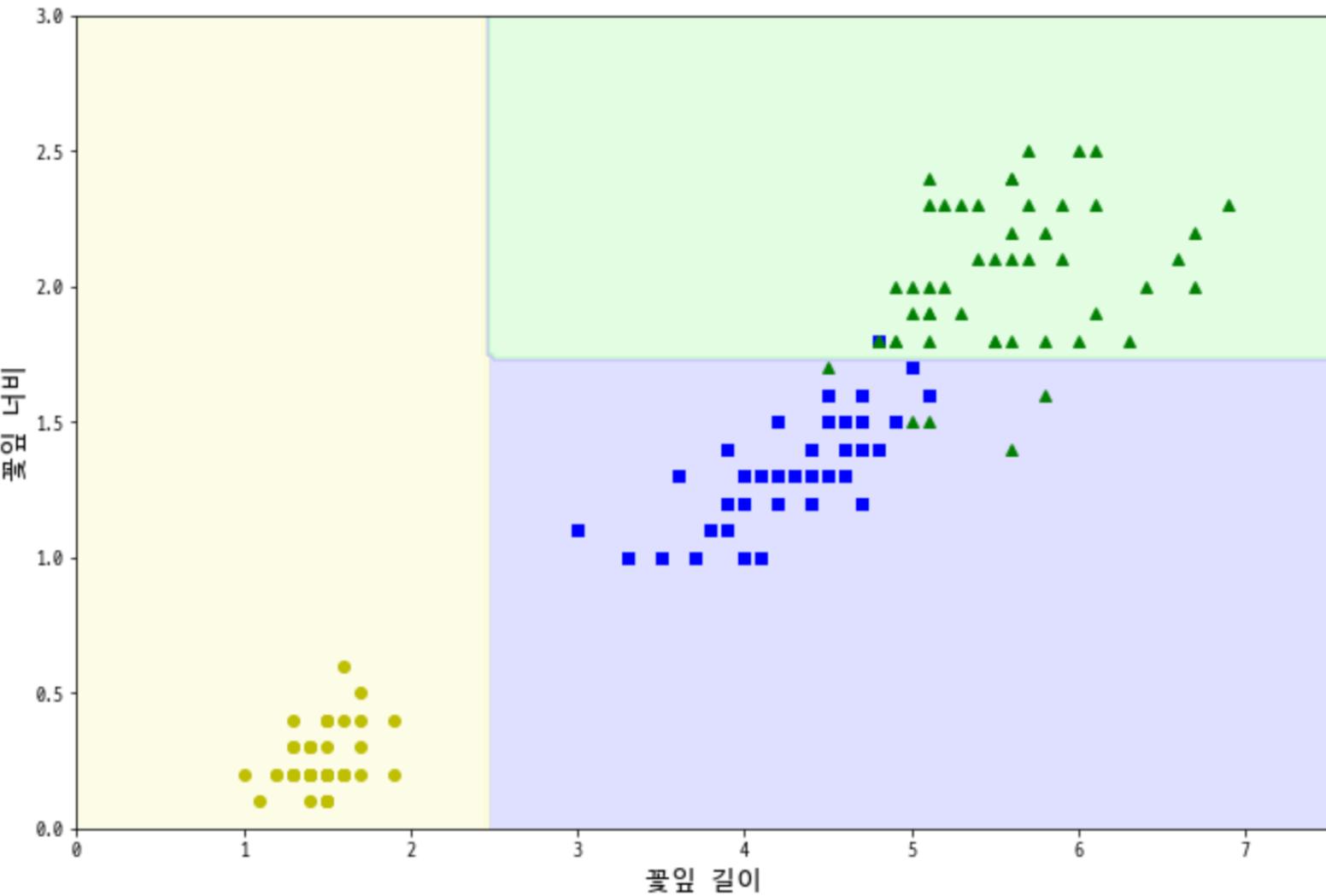
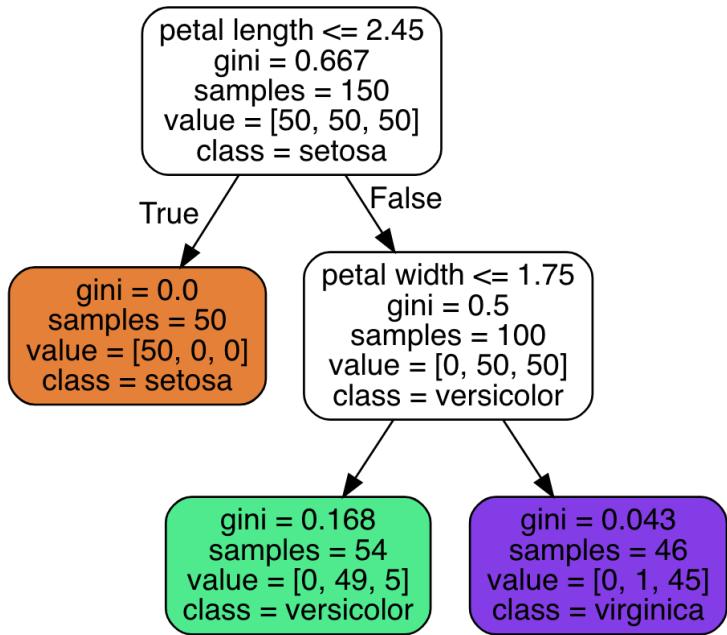
```
array([[1.4, 0.2],  
       [1.4, 0.2],  
       [1.3, 0.2],  
       [1.5, 0.2],
```

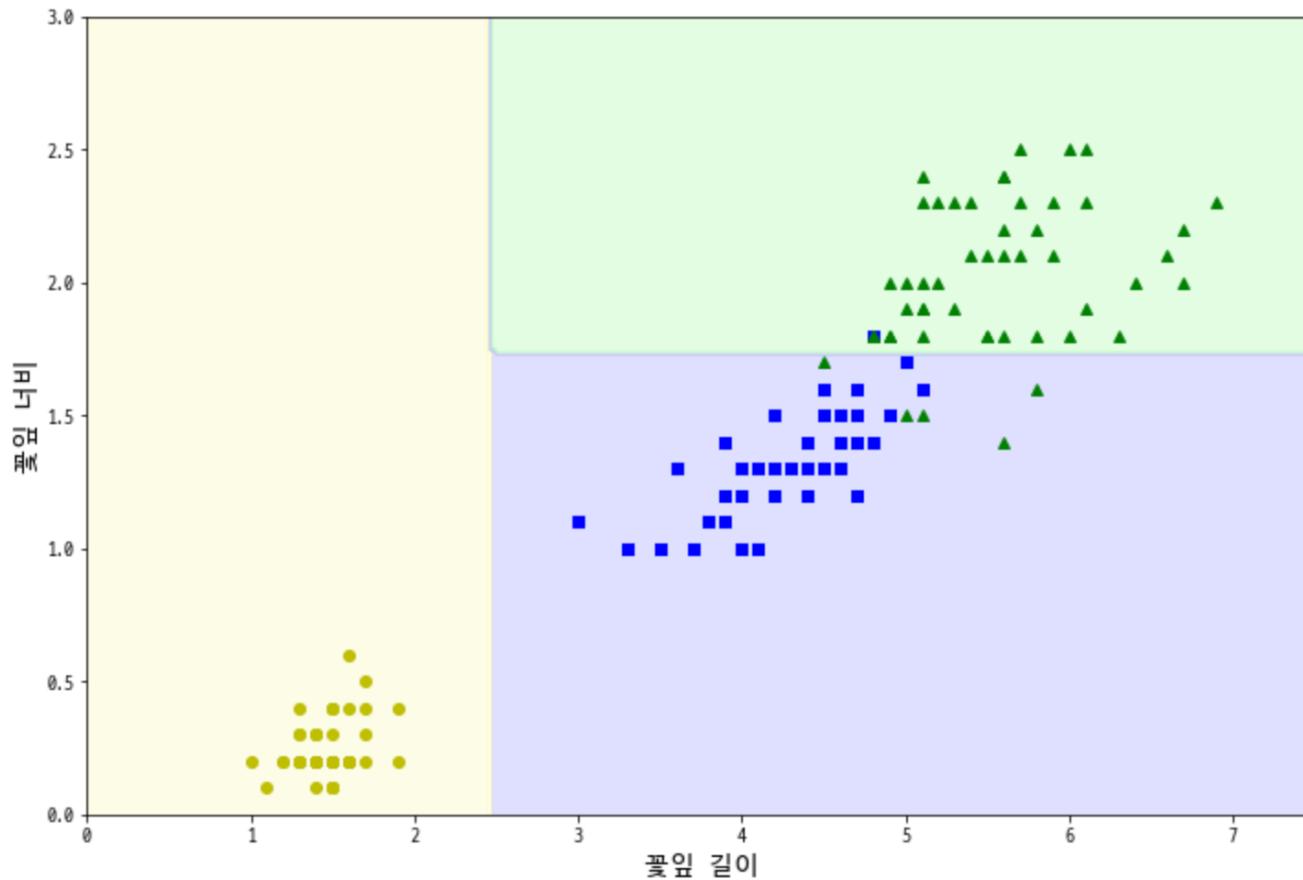
In [31]:

y

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
      0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
      0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,  
      1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,  
      1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2,  
      2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,  
      2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```







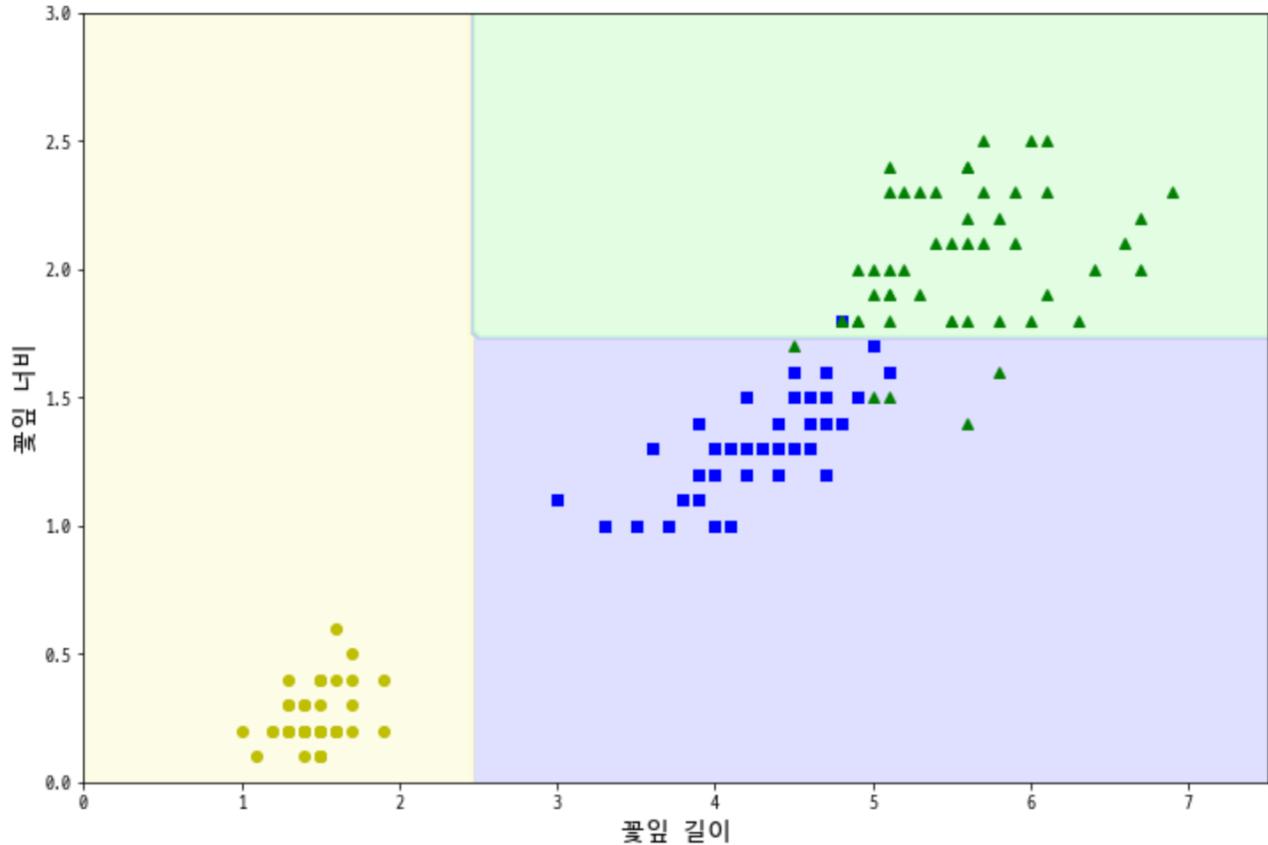
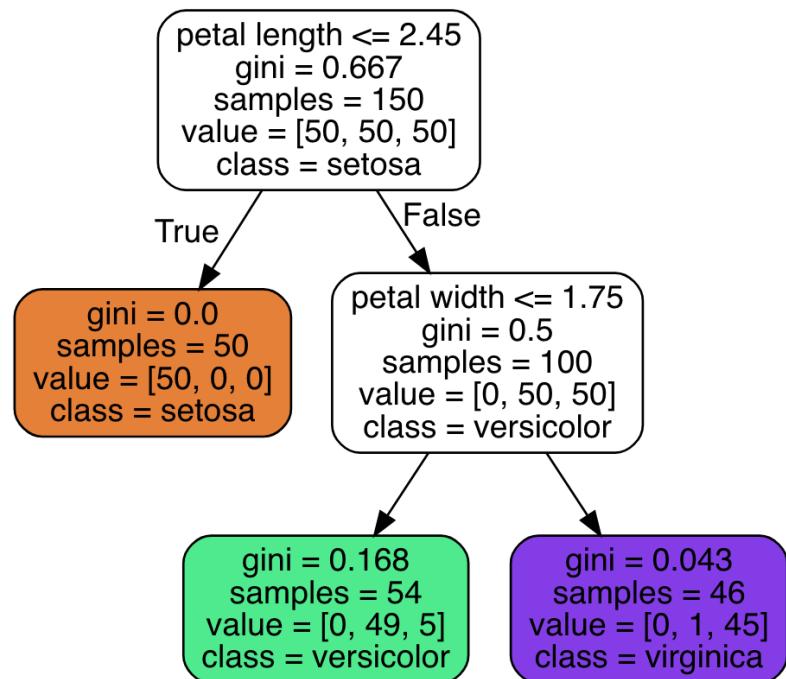
- 저 경계선을 그냥 떼려 잡아야 한다고???

In [32]:

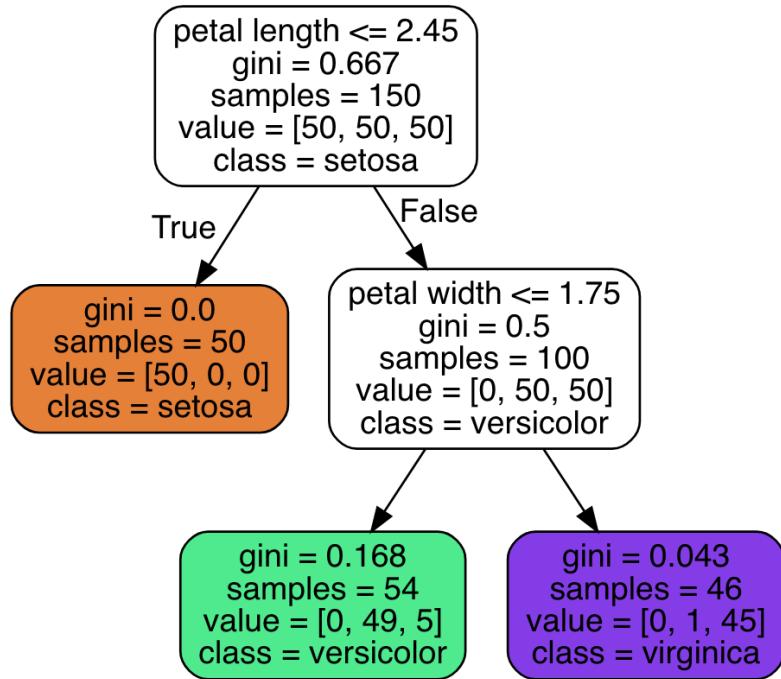
```
tree_clf = DecisionTreeClassifier(max_depth=2, random_state=42)
tree_clf.fit(X, y)
```

```
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=2,
                      max_features=None, max_leaf_nodes=None,
                      min_impurity_decrease=0.0, min_impurity_split=None,
                      min_samples_leaf=1, min_samples_split=2,
                      min_weight_fraction_leaf=0.0, presort=False, random_state=42,
                      splitter='best')
```

- 그럴리가 **scikit learn**에서 준비해 주었다. **DecisionTreeClassifier**



- 그런데... 몇 개로 나눌건지는 설정으로 정해 준다고 하고...
- 경계선을 정하는 건 뭐지??



- **gini** : 불순도
 - 0이면 해당 클래스의 모든 데이터가 참이다

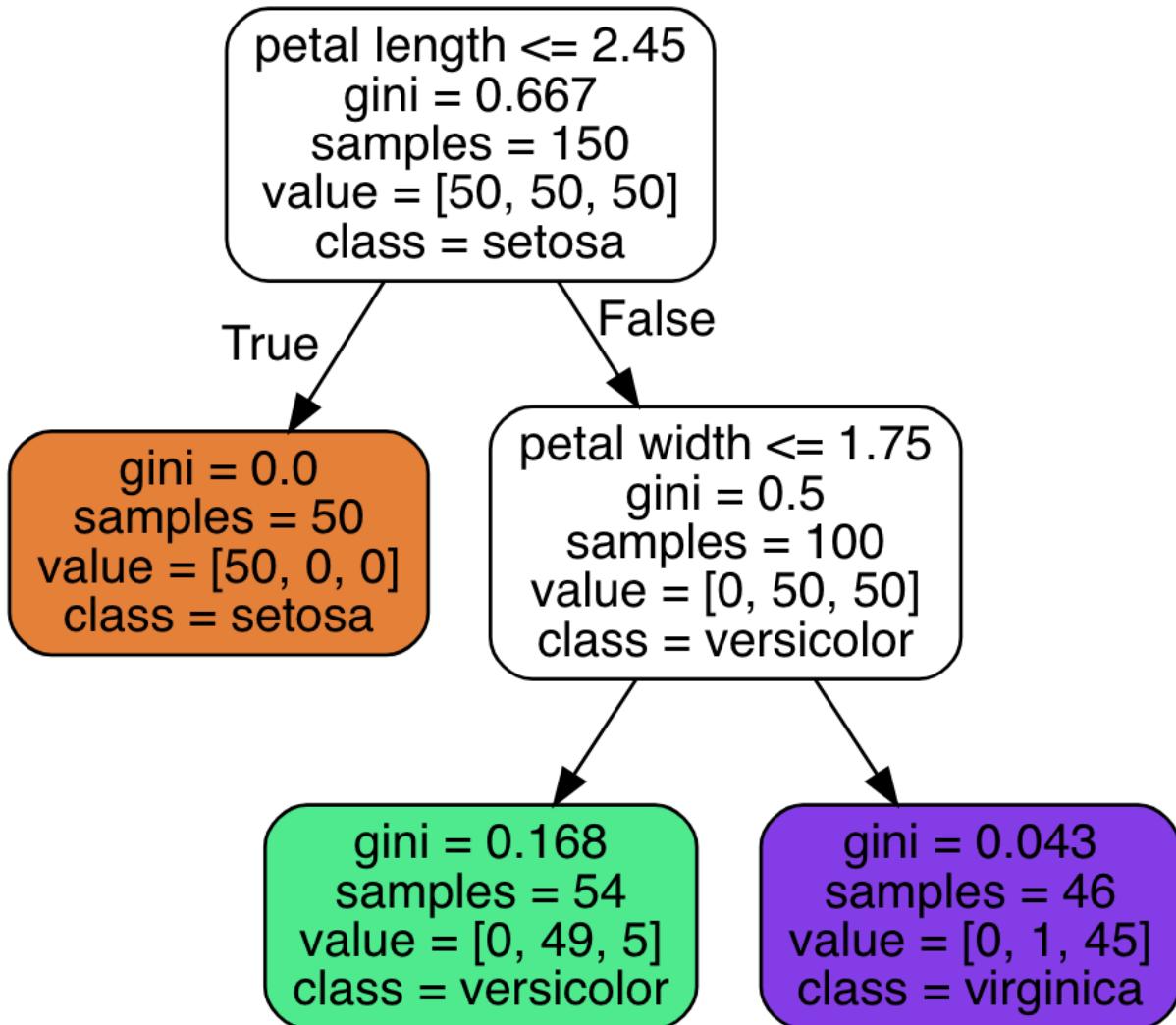
In [33]:

```
from sklearn.tree import export_graphviz

export_graphviz(
    tree_clf,
    out_file="iris_tree.dot",
    feature_names=["petal length", "petal width"],
    class_names=iris.target_names,
    rounded=True,
    filled=True
)
```

In [34]:

```
import graphviz
with open("iris_tree.dot") as f:
    dot_graph = f.read()
dot = graphviz.Source(dot_graph)
dot.format = 'png'
dot.render(filename='iris_tree', directory='images/decision_trees', cleanup=True)
dot
```



In [37]:

```
tree_clf.predict_proba([[5, 1.5]])
```

```
array([[0.          , 0.90740741, 0.09259259]])
```

In [38]:

```
tree_clf.predict([[5, 1.5]])
```

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

- 꽃잎 길이 5, 폭 1.5일때 품종의 종류가 될 확률 **predict_proba**
- 예측 클래스 **predict**