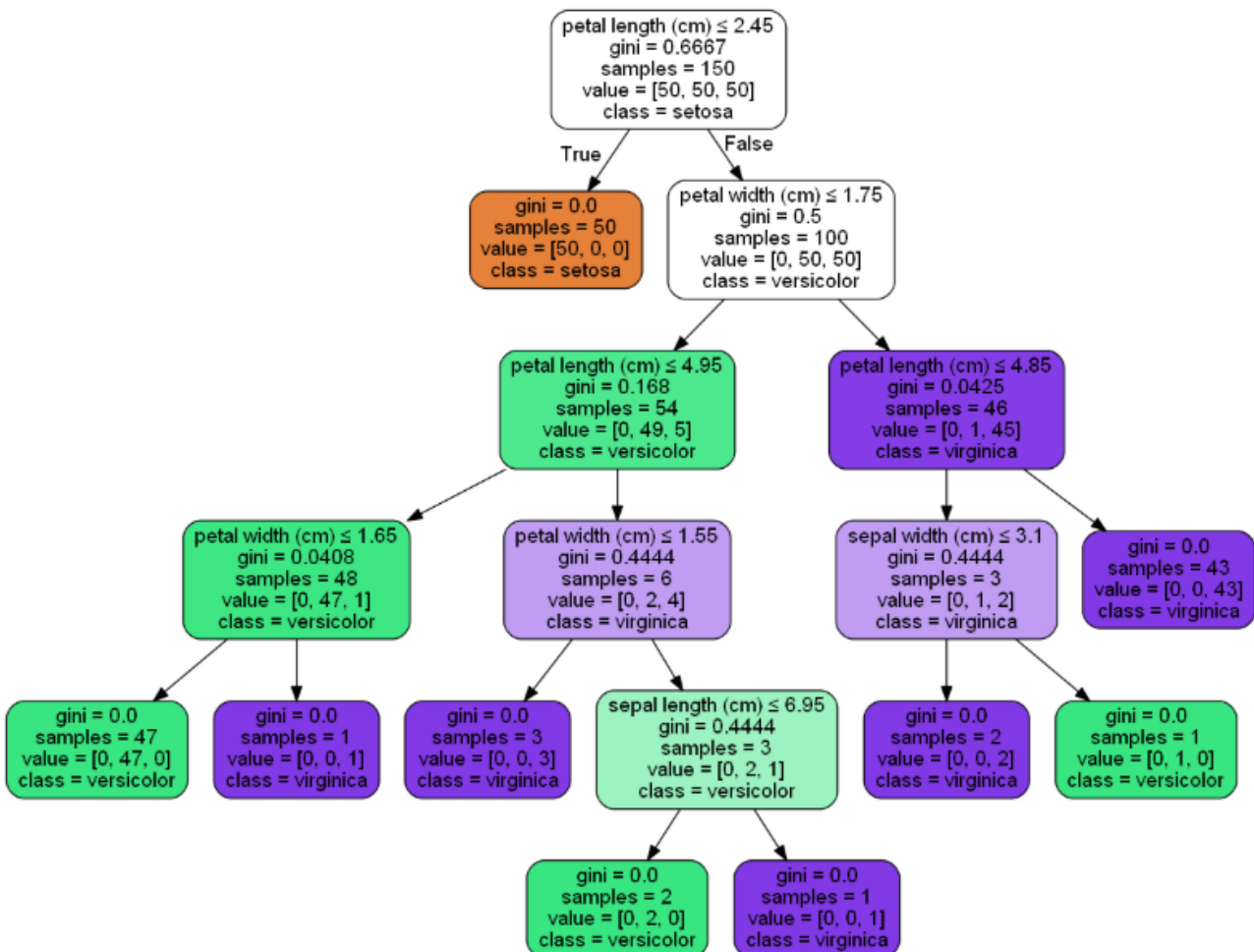


【华子】机器学习009-1-用决策树解决多分类问题

```
from sklearn.datasets import load_iris
import pydotplus
from IPython.display import Image #直接绘制在jupyter中
#训练模型
iris = load_iris()
clf = tree.DecisionTreeClassifier()
clf = clf.fit(iris.data, iris.target)

#绘图
dot_data = tree.export_graphviz(clf, out_file=None,
                                feature_names=iris.feature_names,
                                class_names=iris.target_names,
                                filled=True, rounded=True,
                                special_characters=True)
graph = pydotplus.graph_from_dot_data(dot_data)
Image(graph.create_png())
```



gini系数，判定方法：基尼系数应大于60%，就算好模型。

原理：基尼系数经常用于分类问题，其可以直接从AUC中得到。其公式为： $Gini = 2 * AUC - 1$

如果是内部节点：

第一行是该节点继续二分选择的特征和对应的特征分裂值。

第二行是该节点数据的基尼系数 第三行是训练数据被划分到该节点的样本数。

第四行是训练数据被划分到该节点的样本中，各个类别的样本数量。

第五行是该节点的类别，对应训练数据在该节点的样本中，最多样本的类别

如果是叶子节点：

第一行是该节点数据的基尼系数

第二行是训练数据被划分到该节点的样本数。

第三行是训练数据被划分到该节点的样本中，各个类别的样本数量。

第四行是该节点的类别，对应训练数据在该节点的样本中，最多样本的类别