

# 数据科学之路：01-1课程介绍

## 课程介绍

课程官网: <https://github.com/JnuLi/FAFU-Data-Science-18>

FAFU数据科学研讨班(2018)

data-science fafu Manage topics

14 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

JnuLi 第0次	Latest commit 7522008 4 days ago	
01课程简介	第0次	4 days ago
02数据分析基础	第0次	5 days ago
03机器学习	第1次	7 days ago
README.md	第0次	5 days ago
数据科学导论课程大纲.pdf	第1次	7 days ago

README.md

课程官网

## 授课内容

FAFU数据科学研讨班（2018）主要和大家讨论一些数据科学家需要掌握的基础内容，传授大家如何获取数据、如何分析数据、如何从数据萃取价值、如何应用数据等方面的技能；内容涉及：数据清洗、数据分析、数据可视化、机器学习、推举算法等方面。课程通过**理论授课+实践教学**的方式开展，实践教学通过Python语言来实现。

## 课程安排

- 上课时间：每周的周三、周五的晚上6:30到8:00，周天上午9:00到11:00左右
- 上课地点：宝铃楼201

- 课程安排：预计每周讲一个章节，分三次课来上
- 课程要求：
  - 1、有条件的同学可以带电脑来上课
  - 2、大部分课时都有布置作业，希望大家能按时完成，下次上课是会进行讲解

## 辅助材料

---

### 书籍

---

- 周志华《机器学习》
- 李航《统计学习方法》
- 欧高炎《数据科学导引》

### 课程

---

- 吴恩达《机器学习》
- 林轩田《机器学习基石》和《机器学习技法》
- 李宏毅《机器学习（2017）》

### 网站

---

- [数据嗨客](#)

## 补充

---

### Jupyter上ExecutionTime插件安装

---

在存放文件jupyter\_contrib\_nbextensions-0.4.0-py2.py3-none-any.whl的目录下，按住shift+鼠标右键后，点击在此处打开命令窗口进入命里提示符窗口，依次输入如下指令：

```
pip install jupyter_contrib_nbextensions-0.4.0-py2.py3-none-any.whl
jupyter contrib nbextension install --user
jupyter nbextension enable execute_time/ExecuteTime
```

测试是否安装成功

## 决策树可视化包安装

---

- 1、安装GraphViz。 [GraphViz下载地址](#)
- 2、安装pyparsing和pydot包

```
# pip安装
pip install pyparsing pydot

# conda安装
conda install pyparsing pydot
```

- 3、演示代码

```
from sklearn.datasets import load_iris
import numpy as np
from sklearn import tree
from IPython.display import Image
from sklearn.externals.six import StringIO
import pydot

iris = load_iris()
test_idx = [0,50,100]

# 训练数据
train_target = np.delete(iris.target, test_idx)
train_data = np.delete(iris.data, test_idx, axis=0)

# 测试数据
test_target = iris.target[test_idx]
test_data = iris.data[test_idx]

# 训练分类器
clf = tree.DecisionTreeClassifier()
clf = clf.fit(train_data,train_target)

from sklearn.externals.six import StringIO
import pydot
dot_data = StringIO()
```

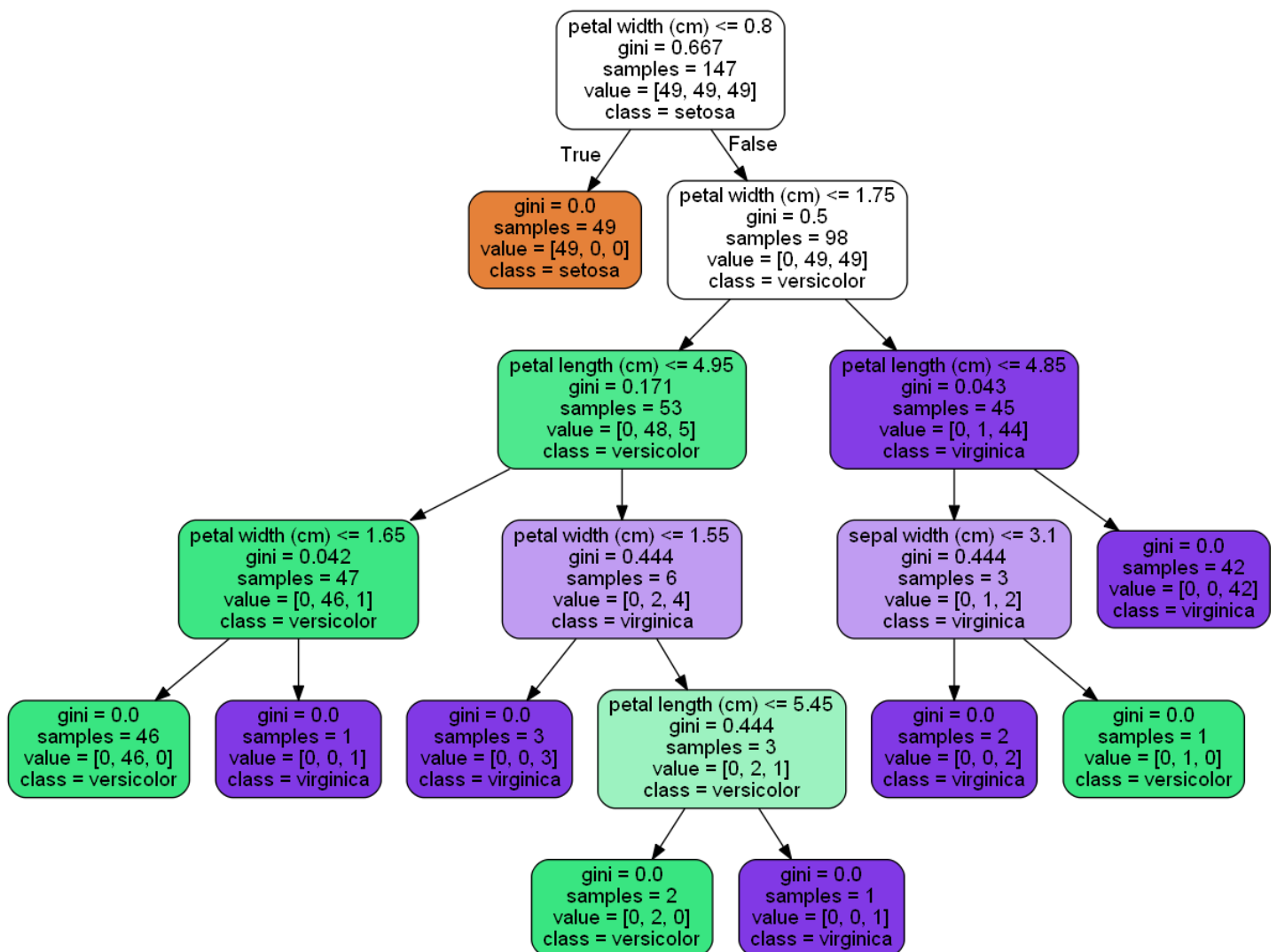
```

tree.export_graphviz(clf, out_file=dot_data,
                    feature_names=iris.feature_names,
                    class_names=iris.target_names,
                    filled=True, rounded=True,
                    impurity=True)

graph = pydot.graph_from_dot_data(dot_data.getvalue())
graph[0].create_png()
graph[0].write_png("pruning_1.png")

```

如果成功安装，会输出如下图形：



决策树可视化