Scikit-Learn学习笔记

官方帮助文档可参考：**2017-11-09 ApacheCN 开源组织**提供**scikit-learn 0.18 中文文档**

机器学习类型

1. 降维

A. 有监督学习 - 有监督神经网络

2. 回归：预测数值

3. 分类：预测类别

B. 无监督学习 - 无监督神经网络

4. 聚类

Machine Learning Project Checklist

This checklist can guide you through your Machine Learning projects. There are eight main steps:

1. Frame the problem and look at the big picture.

2. Get the data.

3. Explore the data to gain insights.

4. Prepare the data to better expose the underlying data patterns to Machine Learning algorithms.

5. Explore many different models and short-list the best ones.

6. Fine-tune your models and combine them into a great solution.

7. Present your solution.

8. Launch, monitor, and maintain your system.

Obviously, you should feel free to adapt this checklist to your needs.

更详尽的流程可参考（./pics/）：Machine Learning Project Checklist

入门实例

```
# -*- coding:utf-8 -*-
from sklearn.svm import SVC
import numpy as np
# 数据准备
rng = np.random.RandomState(0)
X = rng.rand(100, 10)
y = rng.binomial(1, 0.5, 100)
X_test = rng.rand(5, 10)
# 选择SVM模型
clf = SVC()
# 训练带参数优化的模型
clf.set_params(kernel='linear').fit(X, y)  # 线性核
# 拿训练好的模型进行预测
clf.predict(X_test)
# 训练带参数优化的模型
clf.set_params(kernel='rbf').fit(X, y)  # rbf核
# 拿训练好的模型进行预测
clf.predict(X_test)

```