

scikit-learn 教程 - 导航

作者: [Zing](#)

原文链接: <https://ld246.com/article/1463124574611>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

<p>此为目录导航。 </p>
<p>项目里头需要修改某个机器算法，为了方便阅读源码和修改，选择了 python 实现的 sklearn。 <p>
<p>不得不说 sklearn 的文档写得特别好，忍不住翻译一发。 </p>
<p>翻译也是欣赏和学习高手们设计的艺术吧。 </p>
<p>#简介：使 scikit-learn 进行机器学习 </p>

机器学习：问题设定
加载样例数据集
学习和预测
模型持久化
惯例

<hr>
<p>#统计学习程 </p>

统计学习：scikit-learn 中的配置和 estimator 对象
有监督学习：预测高维观测对象
模型选择：选择 estimator 和参数
无监督学习：寻找数据特征
连接所有流程
帮助

<hr>
<p>#文本分析 /a> </p>

设置
加载 “Twenty Newsgroups” 数据集
抽取 text 文件的特征
训练分类器
建立管道
评估测试集的表现
使用网格搜索调整参数
练习 1：语言识别
练习 2：影评情感分析
练习 3：CLI 文本分类应用
路在何方

<hr>
<p>#选择合适的 estimator</p>
<p>通常机器学习最难的一部分是选择合适的 estimator。 </p>
<p>不同的 estimator 适用于不同的数据集和问题。 </p>
<p>在本节中，sklearn 官方文档提供了一个图，可以快速根据你的数据和问题选择合适的 estimator。单击相应的区域还可以获得更具体的内容。 </p>
<p>推荐到官网看看 flowchart ，很有意思，具有简单的普适性</p>
<hr>

<p>#其他的资源，视频和讨论</p>

python 统计学新手

其他的教程

视频

<p>具体参见官方文档

http://scikit-learn.org/stable/presentations.html</p>

<hr>

<blockquote>

<p>注：文档测试模式

上面教程中的示例代码都是以 python 控制台的格式给出。如果你希望方便地在 IPython 中执行这些代码，在 IPython 控制台使用：<code>%doctest_mode</code>。这样就可以直接复制和粘贴代码，而不必要手动地移除 >>></p>

</blockquote>