Python机器学习需要用到的库

Python是一种面向对象的解释型计算机程序设计语言，具有丰富和强大的库，再加上其简单、易学、速度快、开源免费、可移植性、可扩展性以及面向对象的特点，Python成为2017年最受欢迎的最受欢迎的编程语言！

人工智能是当前最热门话题之一，机器学习技术是人工智能实现必备技能，Python编程语言含有最有用的机器学习工具和库，以下是Python开发工程师必知的十大机器学习库！

**一、Scikit-Learn**

在机器学习和数据挖掘的应用中，Scikit-Learn是一个功能强大的Python包，我们可以用它进行分类、特征选择、特征提取和聚集。

**二、Statsmodels**

Statsmodels是另一个聚焦在统计模型上的强大的库，主要用于预测性和探索性分析，拟合线性模型、进行统计分析或者预测性建模，使用 Statsmodels是非常合适的。

**三、PyMC**

PyMC是做“贝叶斯曲线”的工具，其包含贝叶斯模型、统计分布和模型收敛的诊断工具，也包含一些层次模型。

**四、Gensim**

Gensim被称为“人们的主题建模工具”，其焦点是狄利克雷划分及变体，其支持自然语言处理，能将NLP和其他机器学习算法更容易组合在一起，还引用了Google的基于递归神经网络的文本表示法word2vec。

**五、Orange**

Orange是一种带有图形用户界面的库，在分类、聚集和特征选择方法方面，相当齐全，还有交叉验证的方法。

**六、PyMVP**A

PyMVPA是一种统计学习库，包含交叉验证和诊断工具，但没有Scikit-learn全面。

**七、Theano**

Theano是最成熟的深度学习库，它提供了不错的数据结构表示神经网络的层，对线性代数来说很高效，与Numpy的数组类似，很多基于Theano的库都在利用其数据结构，它还支持开箱可用的GPU编程。

**八、PyLearn**

PyLearn是一个基于Theano的库，它给Theano引入了模块化和可配置性，可以通过不同的配置文件来创建神经网络。

**九、Hebel**

Hebel是一个带有GPU支持的神经网络库，可以通过YAML文件决定神经网络的属性，提供了将神级网络和代码友好分离的方式，并快速地运行模型，它是用纯Python编写，是很友好的库，但由于开发不久，就深度和广大而言，还有些匮乏！

**十、Neurolab**

Neurolab是一个API友好的神经网络库，其包含递归神经网络实现的不同变体，如果使用RNN，这个库是同类API中最好的选择之一。

以上是Python开发工程师必知十大机器学习库，除此之外，还有OverFeat、Nolearn以及Decaf等机器学习库，在此就不一一介绍了，感兴趣的可以深入学习一下！