**2020 必学的10大顶级 Python 库**

**python 是最流行和使用最广泛的编程语言之一，它已经取代了业界许多编程语言。python 在开发人员中流行的原因有很多。然而，最重要的一点是它有大量的库供用户使用。  
  
python 的简单性吸引了许多开发人员为机器学习创建新的库。由于有大量的库，python 在机器学习专家中变得非常流行。  
  
所以，这里要介绍的第一个库是 TensorFlow。**

**1.TensorFlow**



什么是 TensorFlow？

如果你目前正在使用 python 进行机器学习项目，那么你可能听说过这一个流行的开源库，那就是 TensorFlow。

这个库是由 Google 与 Brain Team 合作开发的，几乎每一个 Google 的机器学习应用程序都用到了 TensorFlow。

TensorFlow 就像一个计算库，用于编写涉及大量 tensor 操作的新算法。由于神经网络可以很容易地表示为计算图，因此它们可以使用 TensorFlow 作为 tensor 的一系列操作来实现。另外，tensor 是表示数据的 n 维矩阵。

TensorFlow 的特征

1.快速响应的结构

使用 TensorFlow，我们可以很容易地可视化图的每个部分，这在使用 Numpy 或 SciKit 时是做不到的。

2.灵活

TensorFlow 的一个非常重要的特性是，它的操作非常灵活。这意味着它具有模块性，可以让你把希望独立出来的部分分出来

3.容易训练

对于分布式计算来说，它很容易在 CPU 和 GPU 上训练。

4.并行神经网络训练

TensorFlow 提供了管道流，从这个意义上说，你可以训练多个神经网络和多个 GPU，这使得模型在大型系统上非常有效。

5.大型社区

不用说，它是由 Google 开发的，已经有一个庞大的软件工程师团队在不断地改进稳定性。

6.开源

这个机器学习库最好的一个特点是，它是开源的，任何人只要有连接互联网就可以使用它。

**TensorFlow 被用在哪里？**

你每天都在使用 TensorFlow，你使用的 Google Voice Search 或 Google Photos 等应用程序都是使用这个库开发的。

在 TensorFlow 创建的所有库都是用 C 和 C++编写的，但是，它有一个复杂的前端，是用 python 实现的。你的 python 代码将被编译，然后在使用 C 和 C++构建的 TensorFlow 分布式执行引擎上执行。

实际上，TensorFlow 的应用是无限的，这就是它美妙的地方。

**2.Scikit-Learn**



**什么是 Scikit-Learn？**

它是一个与 NumPy 和 SciPy 相关联的 python 库。它被认为是处理复杂数据的最佳库之一。

在这个库中进行了许多修改。其中一个修改是交叉验证特性，它提供了使用多个度量的能力。许多训练方法，如物流回归和最邻近算法，都没有得到什么改善。

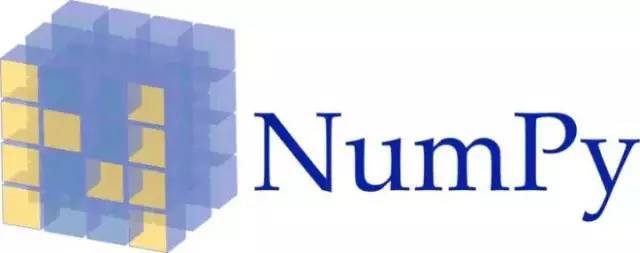
**Scikit-Learn 的特性**

* 交叉验证：有多种方法可以检查不可见数据上受监督模型的准确性。
* 无监督学习算法：同样，在产品中有大量的算法——从聚类、因子分析、主成分分析到无监督神经网络
* 特征提取：用于从图像和文本中提取特征（例如一段文字）

**Scikit Learn 被用在哪里？**

它包含许多实现标准机器学习和数据挖掘任务的算法，如降维、分类、回归、聚类和模型选择。

**3.Numpy**



**什么是 Numpy？**

Numpy 被认为是 python 中最流行的机器学习库之一。

TensorFlow 和其他库在内部使用 Numpy 对 tensor 执行多个操作。数组接口是 Numpy 的最佳和最重要的特性。

**Numpy 的特性**

* 交互性：Numpy 非常容易理解和使用
* 数学性：使复杂的数学实现变得非常简单
* 直观：真正使编码变得容易，掌握概念也很容易
* 大量接口：广泛使用，因此有很多开源贡献者

**Numpy 被用在哪里？**

该接口可用于将图像、声音和其他二进制原始流表示为 n 维实数数组。

机器学习库的实现，拥有 Numpy 的知识对于全栈开发人员来说是很重要的

**4.Keras**



**什么是 Keras？**

Keras 被认为是 python 中最酷的机器学习库之一。它提供了一种更容易表达神经网络的机制。Keras 还为编译模型、处理数据集、图形可视化等提供了一些最佳实用程序。

在后端，Keras 在内部使用 Theano 或 TensorFlow。也可以使用一些最流行的神经网络，如 CNTK。当我们将其与其他机器学习库进行比较时，Keras 的速度相对较慢，因为它使用后端基础设施创建计算图，然后利用它执行操作。Keras 的所有模型都很轻简。

**Keras 的特征**

* 它在 CPU 和 GPU 上都能顺利运行。
* Keras 支持几乎所有的神经网络模型——全连接、卷积、池化、循环、嵌入等。此外，这些模型可以结合起来构建更复杂的模型。
* Keras 本质上是模块化的，具有难以置信的表现力、灵活性和创新性研究的能力。
* Keras 是一个完全基于 python 的框架，它使调试和探索变得容易。

**Keras 被用在哪里？**

你已经在不断地与使用 Keras 构建的产品进行交互—Netflix、Uber、Yelp、Instacart、Zocdoc、Square 和许多其他公司都在使用它。它在初创企业中尤其受欢迎，初创企业将深度学习放在其产品的核心位置。

Keras 包含许多常用的神经网络构建块的实现，例如层、目标、激活函数、优化器和一系列工具，以使图像和文本数据的处理更加容易。

此外，它还提供许多预处理的数据集和预训练的模型，如 MNIST, VGG, Inception, SqueezeNet, ResNet 等。

Keras 也是深度学习研究人员的最爱。大型科学组织，特别是 CERN and NASA 的研究人员尤其偏爱 Keras。

**5.PyTorch**



**什么是 PyTorch？**

PyTorch 是最大的机器学习库，它允许开发人员以 GPU 的加速度执行 tensor 计算，创建动态计算图，并自动计算梯度。除此之外，PyTorch 还提供了丰富的 API 来解决与神经网络相关的应用程序问题。

这个机器学习库是基于 Torch 的，它是一个用 C 语言实现的开源机器库，在 Lua 中进行了封装。

此机器学习库（python）于 2017 年推出，自其问世以来，该库越来越受欢迎，并吸引了越来越多的机器学习开发人员。

**PyTorch 的特性**

* 端到端 Hybrid

一种新的混合前端，提供了易于使用和具有灵活性的 Eager Mode，同时为了速度，无缝过渡到 graph mode，在 C++运行环境中非常实用。

* 分布式训练

利用本地支持异步执行集体操作和点对点通信（Python 和 C++），优化研究和生产中的性能。

* python 优先

PyTorch 不是一个将 python 绑定到 C++框架的工具。它的构建是为了深入集成到 python 中，以便可以与流行的库和包（如 Cython 和 Numba）一起使用。

* 库和工具

一个由研究人员和开发人员组成的活跃社区已经建立了一个丰富的工具和库的生态系统，用于扩展 PyTorch 并支持从计算机视觉到强化学习等领域的开发。

**PyTorch 被用在哪里？**

PyTorch 主要用于自然语言处理等领域的应用程序。

它主要是由 Facebook 的人工智能研究小组开发的，Uber 的概率编程软件「Pyro」就建立在它的基础之上。

PyTorch 在很多方面都优于 TensorFlow，最近它得到了很多关注。

**6.LightGBM**



**什么是 LightGBM？**

梯度增强是最好的和最流行的机器学习（ML）库之一，它可以帮助开发人员使用重新定义的基本模型，即决策树来构建新的算法。因此，有专门的库可以快速有效地实现这种方法。

这些库包括 LightGBM, XGBoost 和 CatBoost。这些库之间存在相互竞争的关系，它们都有助于解决常见问题，可以以几乎相似的方式使用。

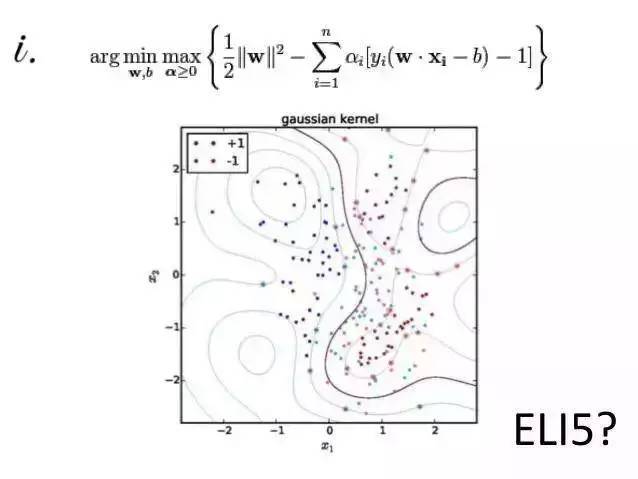
**LightGBM 的特点**

* 计算速度快，生产效率高。
* 直观，易于使用。
* 比其他许多深度学习库更快地训练。
* 在遇到 NaN 值和其他规范值时不会产生错误。

**LightGBM 被用在哪里？**

这个库提供了高度可扩展、优化和快速的梯度增强实现，这使得它在机器学习开发人员中很受欢迎。大多数机器学习全栈开发人员通过使用这些算法赢得了机器学习竞赛。

**7.Eli5**



**什么是 Eli5？**

通常，机器学习模型预测的结果并不准确，python 内置的机器学习库 Eli5 有助于克服这一挑战。它是可视化和调试所有机器学习模型的组合，并跟踪算法的所有工作步骤。

**Eli5 的特点**

此外，Eli5 还支持其他库，包括 xgboost、lightning、scikit-learn 和 sklearn-crfsite。所有上述库中额每一个都可以执行不同的任务。

**Eli5 被用在哪里？**

* 在短时间内需要进行大量计算的数学应用
* Eli5 在和其他 Python 包存在依赖关系的情况下发挥着至关重要的作用
* 在各个领域的传统应用程序实现新方法

**8.SciPy**



**什么是 SciPy？**

SciPy 是一个面向应用程序开发人员和工程师的机器学习库。但是，你仍然需要知道 SciPy 库和 SciPy 堆栈之间的区别。SciPy 库包含用于优化、线性代数、集成和统计的模块。

**SciPy 的特点**

* SciPy 库的主要特点是它是使用 Numpy 开发的，它的数组充分利用了 Numpy。
* 此外，SciPy 还使用其特定的子模块提供了所有有效的数值程序，如优化、数值积分和许多其他程序。
* 所有 SciPy 子模块中的所有功能都有具体的文档注释。

**SciPy 被用在哪里？**

SciPy 是一个使用 Numpy 来解数学函数的库。SciPy 使用 Numpy 数组作为基本数据结构，并附带用于科学编程中各种常用任务的模块。

SciPy 可以轻松地处理线性代数、积分（微积分）、常微分方程求解和信号处理等任务。

**9.Theano**



**什么是 Theano？**

Theano 是一个用于计算多维数组的计算框架机器学习库。它的工作原理与 TensorFlow 相似，但不如 TensorFlow 有效，因为它无法适应生产环境。

此外，Theano 也可以在与 TensorFlow 类似的分布式或并行环境中使用。

**Theano 的特点**

* 与 Numpy 紧密集成：能够在无编译函数中使用完整的 Numpy 数组
* 高效地使用 GPU：比 CPU 执行数据密集型计算要快得多
* 有效的符号区分：Theano 为具有一个或多个输入的函数求导数
* 速度和稳定性优化：即使在 x 非常小的情况下，也能求出 log（1+x）的正确答案。这只是一个可以证明 Theano 稳定性的例子
* 动态 C 代码生成：比以前更快地评估表达式，从而大大提高效率
* 广泛的单元测试和自验证：检测和诊断模型中多种类型的歧义和错误

**Theano 被用在哪里？**

Theano 表达式的实际语法是符号化的，这对于习惯于常规软件开发的初学者来说是很不方便的。具体来说，表达式是以抽象的方式定义、编译的，然后直接用于计算。

它是专门为处理深度学习使用的大型神经网络算法所需的计算而设计的。它是同类库中最早的一个（在 2007 年就开始开发了），被认为是深度学习研究和开发的行业标准。

Theano 目前正被用于多个神经网络项目中，而且随着时间的推移，Theano 的普及率也在不断提高。

**10.Pandas**



**什么是 Pandas？**

Pandas 是 Python 中的一个机器学习库，它提供高级的数据结构和各种各样的分析工具。这个库的一个重要特性是能够使用一个或两个命令转换复杂的数据操作。Pandas 有许多内置的分组、数据组合、过滤和时间序列功能的函数。

**Pandas 的特征**

Pandas 确保了整个数据处理的过程更加容易。对诸如重索引、迭代、排序、聚合、连接和可视化等操作的支持是 Pandas 的特色亮点之一。

**Pandas 被用在哪里？**

目前，Pandas 库的版本较少，其中包括数百个新功能、错误修复、增强和 API 更改。Pandas 的改进在于它能够对数据进行分组和排序，为使用的方法选择最适合的输出，并为执行自定义类型的操作提供支持。

当使用 Pandas 的时候，数据分析占了很大的比重。但是，当与其他库和工具一起使用时，Pandas 确保了高性能和良好的灵活性。

总结

python 中的 10 大顶级 机器学习库的介绍就到这里啦，希望本文能够帮助你开始学习 python 中可用的库。