

shiter编写程序的艺术

大数据生态圈，计算机视觉，机器学习，高端技术的爱好者，话不多说，上代码！！...

RSS订阅

原 机器学习项目到底怎么做？站在巨人的肩膀上！-----资料分享

2017年10月22日 16:48:24

阅读数：1655

很久没有更新图形图像处理方面的博客了，最近在培训数据发掘方面的技术，就把学到的东西和大家分享一下。机器学习的项目到底怎么做呢？具体如何和业务结合落地，我们一起来慢慢探索这整个的生态圈。

1. 压箱底的资料

还有一些平时收集的压箱底的资料拿出来和大家分享一下：

1.1 IPOL —经典计算机视觉算法的c实现

http://www.ipol.im/?utm_source=doi



IPOL Journal · Image Processing On Line

[HOME](#) · [ABOUT](#) · [ARTICLES](#) · [PREPRINTS](#) · [WORKSHOPS](#) · [NEWS](#) · [SEARCH](#)

IPOL is a research journal of image processing and image analysis which emphasizes the role of mathematics as a source for algorithm design and the reproducibility of the research. Each article contains a text on an algorithm and its source code, with an online demonstration facility and an archive of experiments. Text and source code are peer-reviewed and the demonstration is controlled. IPOL is an Open Science and Reproducible Research journal.

[Editorial Policy](#) [Editorial Board](#) [Submit an Article](#) [Follow IPOL](#)[Index](#) · [Articles 2011 2012 2013 2014 2015 2016 2017](#) · [Preprints](#) · [News](#) · [Citation score](#)

Latest Articles

- **An Algorithm for Gaussian Texture Inpainting**
2017-10-09 · Bruno Galerne, Arthur Leclaire
- **2D Filtering of Curvilinear Structures by Ranking the Orientation Responses of Path Operators (RORPO)**
2017-10-01 · Odyssee Merveille, Benoît Naegel, Hugues Talbot, Laurent Najman, Nicolas Passat
- **Hyperspectral Image Classification Using Graph Clustering Methods**
2017-08-19 · Zhaoyi Meng, Ekaterina Merkurjev, Alice Koniges, Andrea L. Bertozzi
- **The Image Curvature Microscope: Accurate Curvature Computation at Subpixel Resolution**
2017-07-28 · Adina Ciomaga, Pascal Monasse, Jean-Michel Morel
- **Iterative Hough Transform for Line Detection in 3D Point Clouds**
2017-07-19 · Christoph Dalitz, Tilman Schramke, Manuel Jeltsch
- **Realistic Film Grain Rendering**
2017-07-18 · Alasdair Newson, Noura Faraj, Bruno Galerne, Julie Delon

Series and Special Issues

- [Special Issue on Discrete Geometry \(DGI 2011\)](#)
- [SIIMS Companion Papers](#)

Topics

<http://blog.csdn.net/wangyaninglm>1.2 <https://www.codecademy.com/> —在线编程自学成才

我的python就是在这个网站自学的，基本上把python的基本数据结构，list，dict等等介绍了一遍，只要一周左右甚至更短的时间就可以基本掌握一门全新的语言

BY SUBJECT

Web Development

Partnerships

Programming

Data Science

BY LANGUAGE

HTML & CSS

Python

JavaScript

Java

SQL

Bash/Shell

Ruby

Python

Python is a general-purpose, versatile and modern programming language. It's great as a first language because it is concise and easy to read. Use it for everything from web development to software development and scientific applications.

COURSES

These courses are designed to learn a specific language or framework. Projects and quizzes content are available for Pro members.

Learn Python

Learn to program in Python, a powerful language used by sites like YouTube and

10 hours

<http://blog.csdn.net/wangyaninglm>

1.3 在线绘制框图——没有visio的最好选择

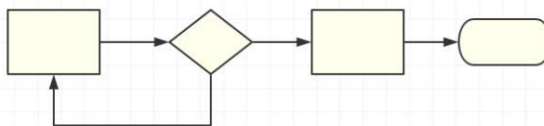
<https://www.processon.com/>

ProcessOn

推

免费在线作图，实时协作

ProcessOn 支持流程图、思维导图、原型图、UML、网络拓扑图等



高效易用、轻松绘制

<http://blog.csdn.net/wangyaninglm>

如果没有visio这是最好的选择！

1.4 一些大牛的博客

刘未鹏

<http://mindhacks.cn/>

<http://mindhacks.cn/2011/11/04/how-to-interview-a-person-for-two-years/>

「码字计划」：拿万元写作基金！ Python薪资多少？

python机器学习

登录

注册

×

廖雪峰的python教程

<https://www.liaoxuefeng.com/wiki/0014316089557264a6b348958f449949df42a6d3a2e542c000>


1.5 写技术博客的选择

在csdn耕耘了8年有余，一直很喜欢这里，然而不知道为何身边的朋友高手就是github，stackoverflow。学习过程中不时记录总结的习惯非常重要，正所谓温故而知新。

后面我也准备尝试一些其他的平台

- 知乎：<https://www.zhihu.com/people/wynshiter/activities>
- <https://cn.wordpress.org/>
- 简书：
- <https://pages.github.com/>

1.6 机器学习算法的c++ sdk (提供更多选择)



Dlib C++ Library

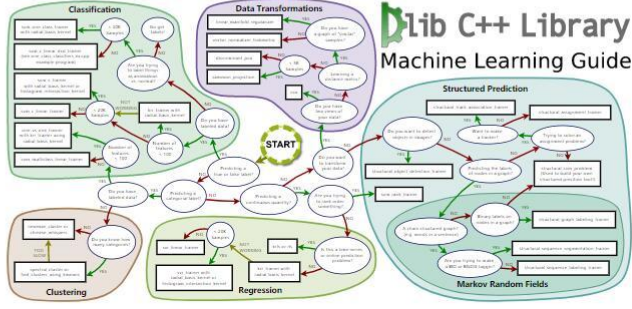
The Library

- Algorithms
- API Wrappers
- Bayesian Nets
- Compression
- Containers
- Graph Tools
- Image Processing
- Linear Algebra
- Machine Learning
- Metaprogramming
- Miscellaneous
- Networking
- Optimization
- Parsing

Help/Info

- Dlib Blog
- Examples: C++
- Examples: Python
- FAQ
- Home
- How to compile
- How to contribute
- Index
- Introduction
- License
- Python API

Machine Learning



Dlib C++ Library Machine Learning Guide

Dlib contains a wide range of machine learning algorithms. All designed to be highly modular, quick to execute, and simple to use via a clean and modern C++ API. It is used in a wide range of applications including robotics, embedded devices, mobile phones, and large high performance computing environments. If you use dlib in your research please cite:

Davis E. King. [Dlib-ml: A Machine Learning Toolkit.](#)

Primary Algorithms

Binary Classification

- `svm_c_ekm_trainer`
- `svm_c_linear_dcd_trainer`
- `svm_c_linear_trainer`
- `svm_c_trainer`
- `svm_nu_trainer`
- `svm_pegasos`
- `train_probabilistic_decision_function`

Multiclass Classification

- `one_vs_all_trainer`
- `one_vs_one_trainer`
- `svm_multiclass_linear_trainer`

Regression

- `krrs`
- `krr_trainer`
- `mlp`
- `rbf_network_trainer`
- `rls`
- `rr_trainer`
- `rvm_regression_trainer`
- `svr_linear_trainer`
- `svr_trainer`

<http://blog.csdn.net/wangyaninglm>

Dlib是一个使用现代C++技术编写的跨平台的通用库，遵守Boost Software licence.它包含机器学习算法和工具，用于在C++中创建复杂的软件来解决现实问题。它在工业和学术界被广泛应用于各种领域，包括机器人，嵌入式设备，移动电话和大型高性能计算环境。

主要特点如下：

- 1.完善的文档：每个类每个函数都有详细的文档，并且提供了大量的示例代码，如果你发现文档描述不清晰或者没有文档，告诉作者，作者会立刻添加。
- 2.可移植代码：代码符合ISO C++标准，不需要第三方库支持，支持win32、Linux、Mac OS X、Solaris、HPUX、BSDs 和 POSIX 系统
- 3.线程支持：提供简单的可移植的线程API
- 4.网络支持：提供简单的可移植的Socket API和一个简单的Http服务器
- 5.图形用户界面：提供线程安全的GUI API
- 6.数值算法：矩阵、大整数、随机数运算等
- 7.机器学习算法：
- 8.图形模型算法：
- 9.图像处理：支持读写Windows BMP文件，不同类型色彩转换
- 10.数据压缩和完整性算法：CRC32、Md5、不同形式的PPM算法
- 11.测试：线程安全的日志类和模块化的单元测试框架以及各种测试assert支持

「码字计划」：拿万元写作基金！ Python薪资多少？

python机器学习

登录

注册

×

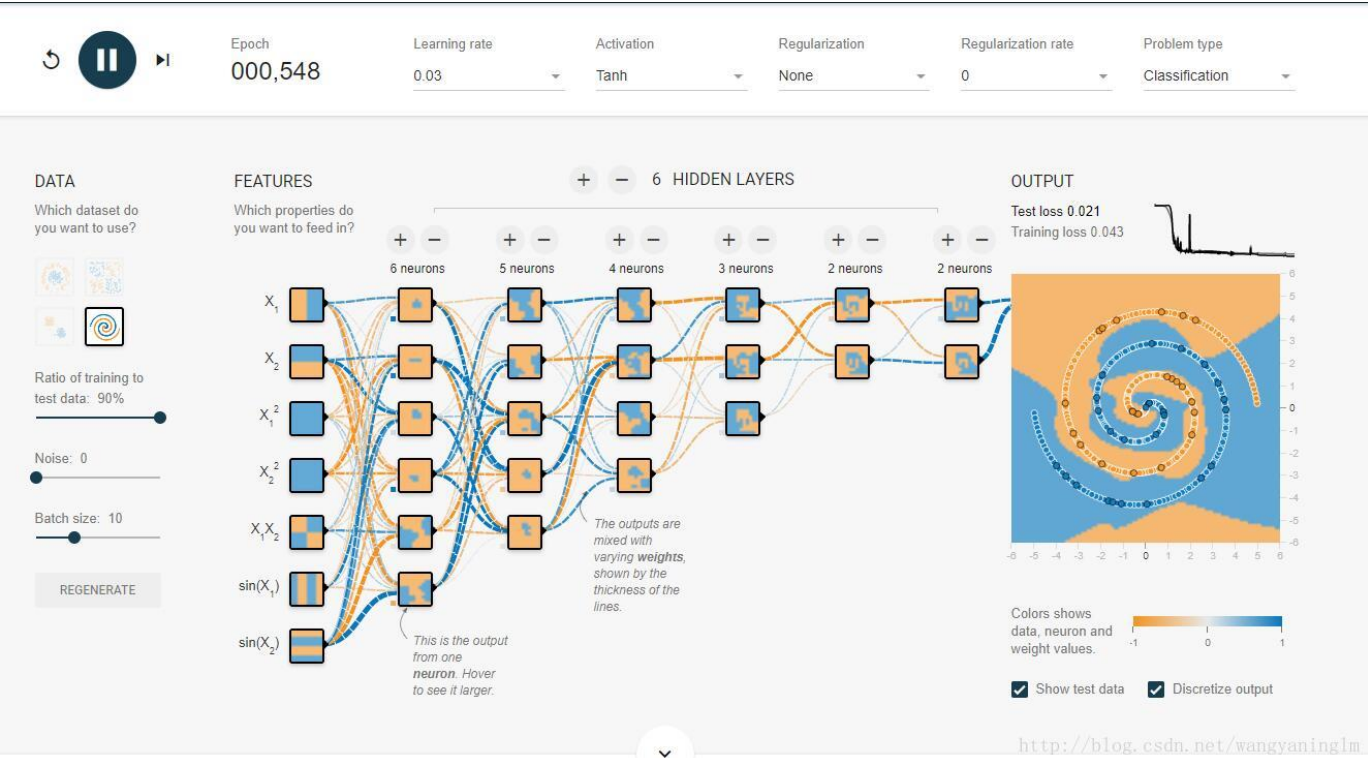
参考网页：

- 2017年最牛逼的五个机器学习项目
- <https://www.kdnuggets.com/2017/01/five-machine-learning-projects-cant-overlook-january.html>
- 35个最牛逼的机器学习项目
- <https://mp.weixin.qq.com/s/zBaOHSMqC7v7dML9AWPLiA>
- 使用dlib 的python接口实现换脸
- <http://python.jobbole.com/82546/>

1.7 可视化感受机器学习的整个过程

1.神经网络的训练：<http://playground.tensorflow.org>

我们选择一个数据为**非线性切分**的复杂例子

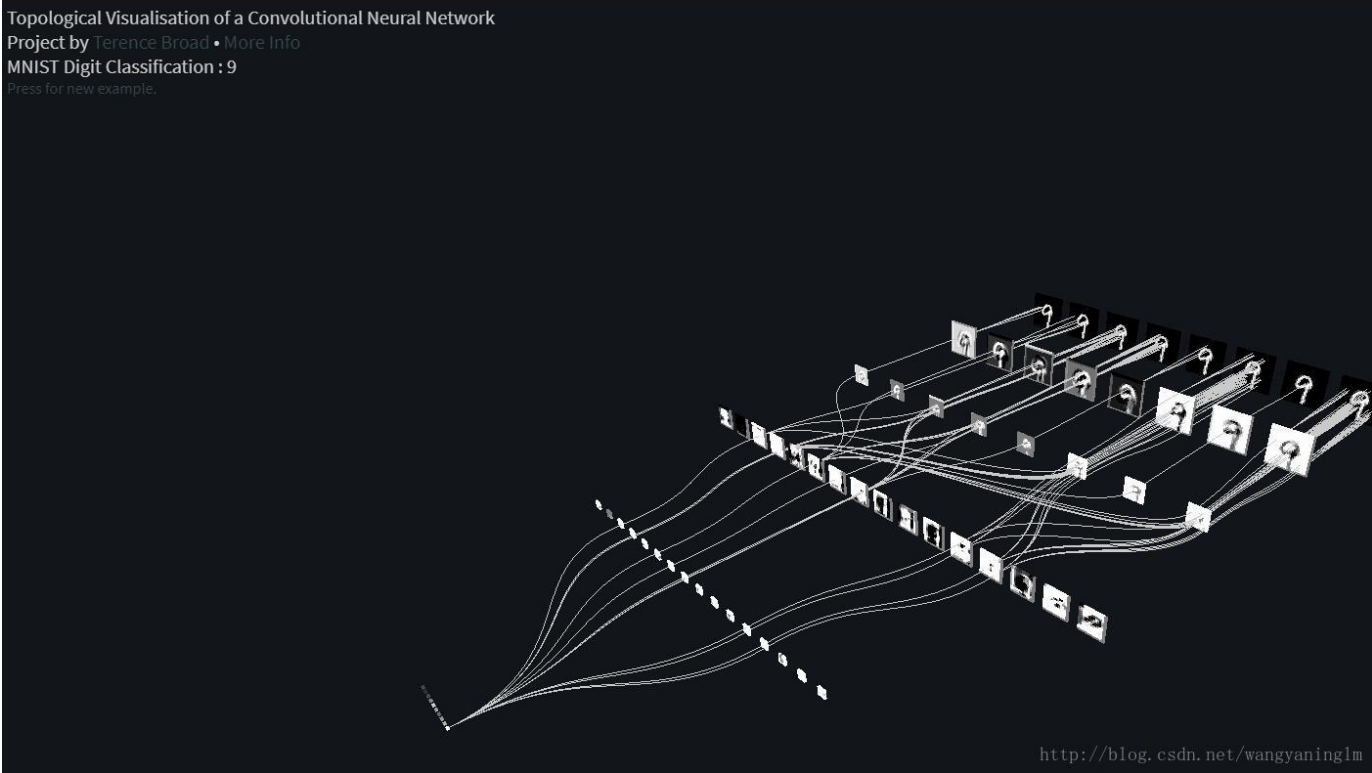


2.两层简单神经网络演示非线性切分

<http://cs.stanford.edu/people/karpathy/convnetjs/demo/classify2d.html>

3.神经网络进行手写字符识别

<http://terencebroad.com/convnetvis/vis.html>



2. Transwarp 机器学习培训

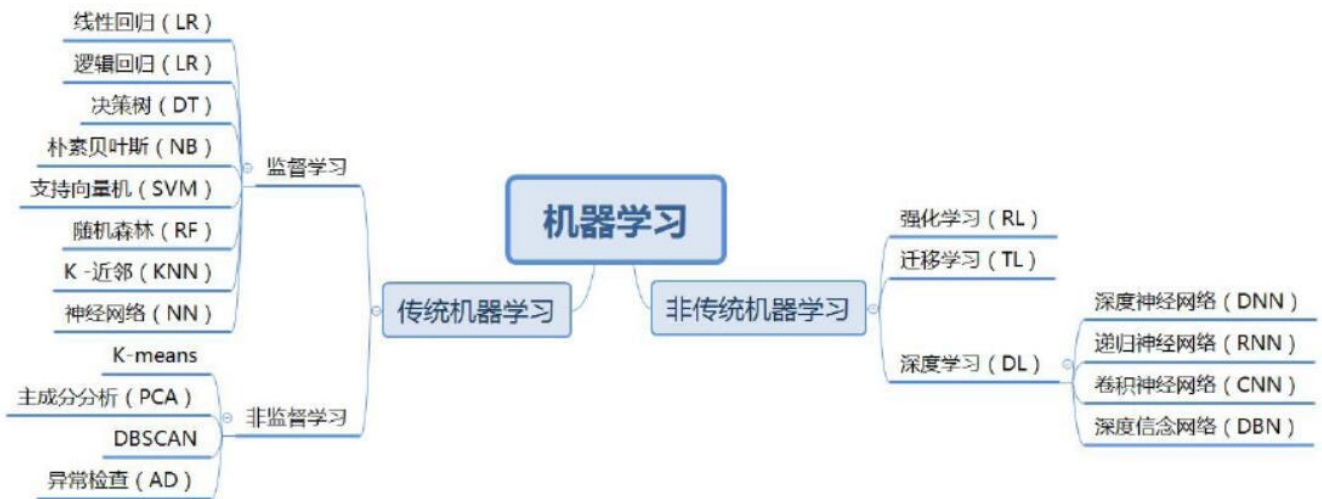
转型机器学习方向的过程是痛苦的。最近在上海花了一周时间参加transwarp的数据分析师培训，这是我时隔7年之后再次踏上魔都的土地。上次来这里还是7年前来看上海世博会，不同的是此处上海之行是由北京启程。我乘坐的是最早一班复兴号列车，由于很多外国人都在新奇 的拍照，这一路风驰电掣的路过祖国的大好河山，让我也怀揣着满满的民族自豪感惊异于祖国 的发展速度。2010年那会来上海一趟多难呀，尤其要买个卧铺，真是难上加难。

非常感谢单位领导给予的宝贵培训机会，之前说实话并未有全面系统的学过机器学习内容。但最重要的还是不知道：真实，工业级，业务上究竟怎么开展机器学习与业务的结合工作。这次培训基本给了我答案。transwarp 通过

推荐其支持拖拉拽的机器学习产品sophon，让我直观的感受了整套机器学习工具平台的使用过程，以及机器学习模型的建模套路。其中之前我一直不太注重的有以下两点：

- 1.特征工程，归一化，字符串索引
- 2.评价指标，roc，方差和等

2.1 机器学习的算法

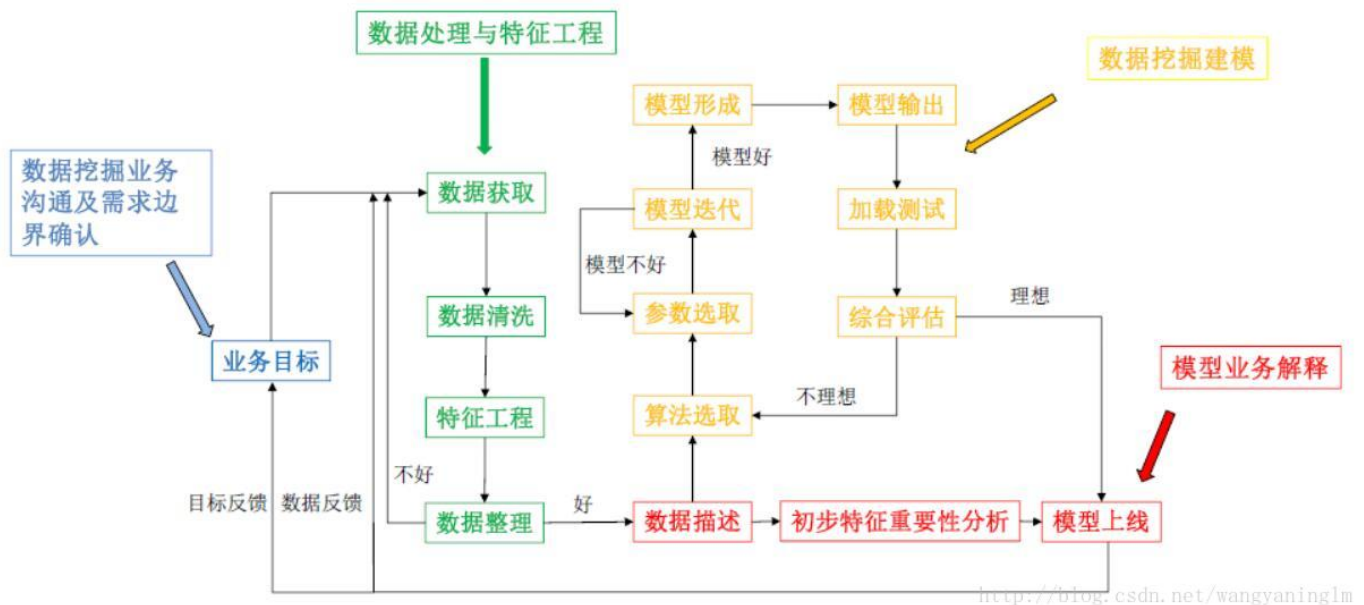


<http://blog.csdn.net/wangyaninglm>

2.2 到底如何衡量业务是否需要机器学习？

1. 业务问题是否适用机器学习算法？
2. 如何选择模型
3. 设计开发节奏
4. 最终产品的检验

2.3 完整的数据发掘建模流程



<http://blog.csdn.net/wangyaninglm>

2.4 特征工程

特征工程是机器学习的决定性因素，**是机器学习成功的关键**

“数据和特征决定了机器学习的上限，而模型和算法只是逼近这个上限而已”

纵观Kaggle、KDD，阿里天池等国内外大大小小的比赛，每个竞赛的冠军其实并没有用到很深的算法，大多数都是在特征工程这个环节做出了出色的工作，然后使用一些常见的算法，比如Linear Regression（线性回归），就能得到出色的

结果

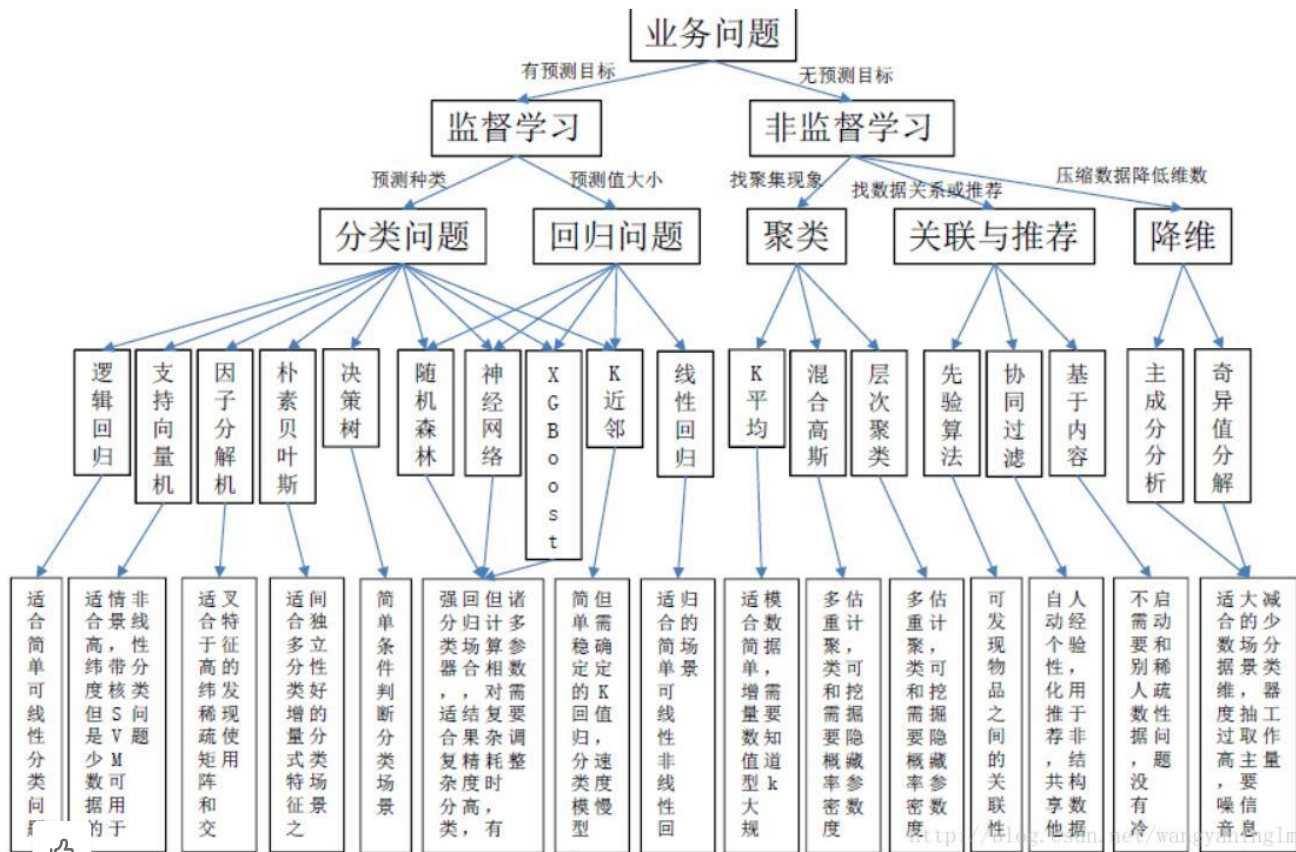
「码字计划」：拿万元写作基金！ Python薪资多少？

python机器学习

登录

注册

×



3

最近还看到号上面一些好的文章，整取领悟以后分享出来

写评论

2.5 可托拉拽机器学习产品

目录

感觉以下产品有点类似

收藏

KNIME <https://www.knime.com/>

收藏

对于机器学习数据科学的初学者来说，最大的挑战之一是需要同时学习太多知识，特别是如果你不知道如何编码。你需要快速地适应线性代数、统计以及机器学习概念，并学习如何编码它们，对于新用户来说，这可能会有点难以承受。

快问

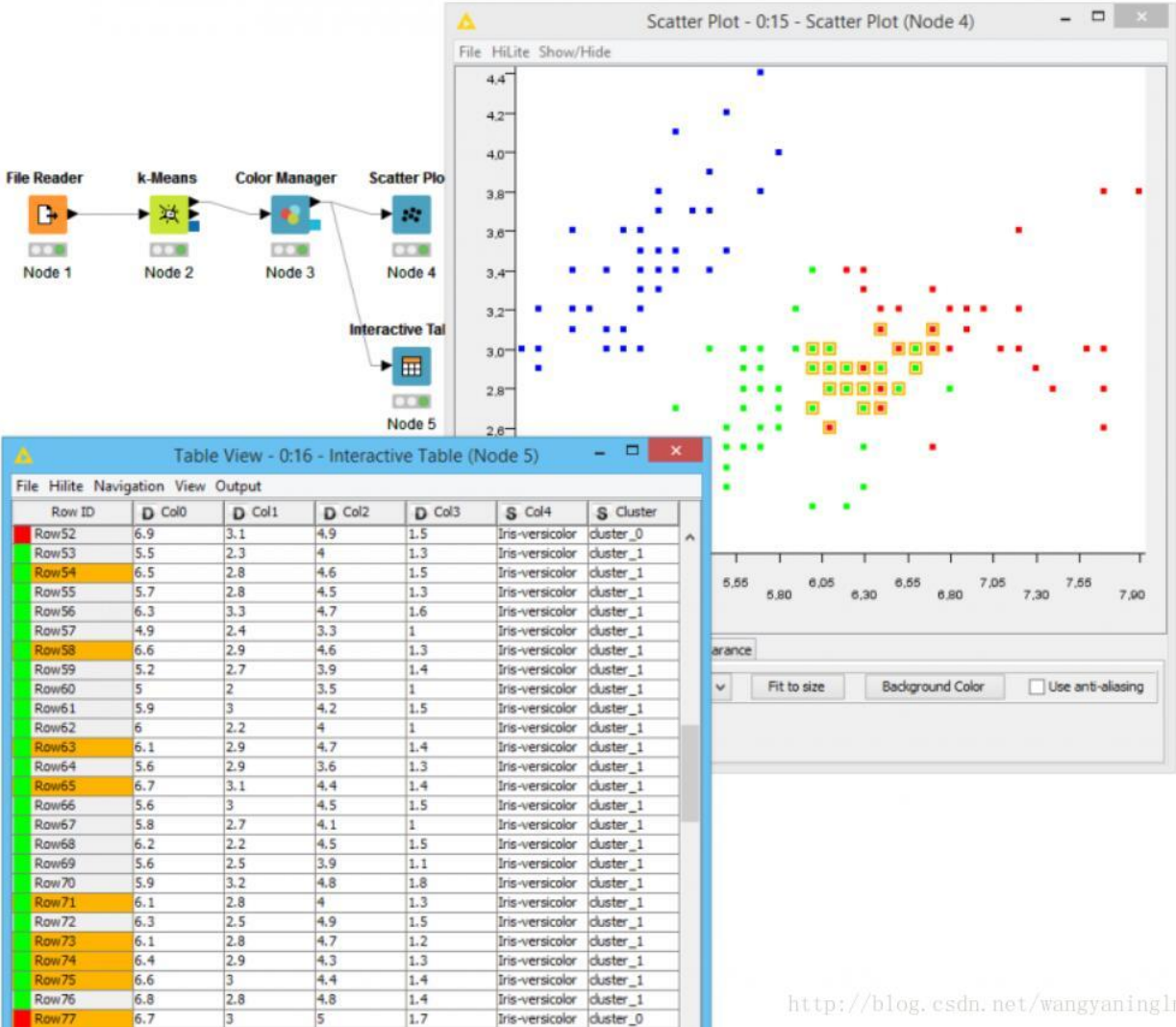
如果你没有编程背景并且发现很难学习下去，这时你可以用一个GUI驱动的工具来学习数据科学。当你刚开始学习的时候，可以集中精力学习实际的项目。一旦适应了基本的概念，你就可以在以后慢慢学习如何编写代码。

微信

微博

QQ

在今天的文章中，我将介绍一个基于GUI的工具：KNIME

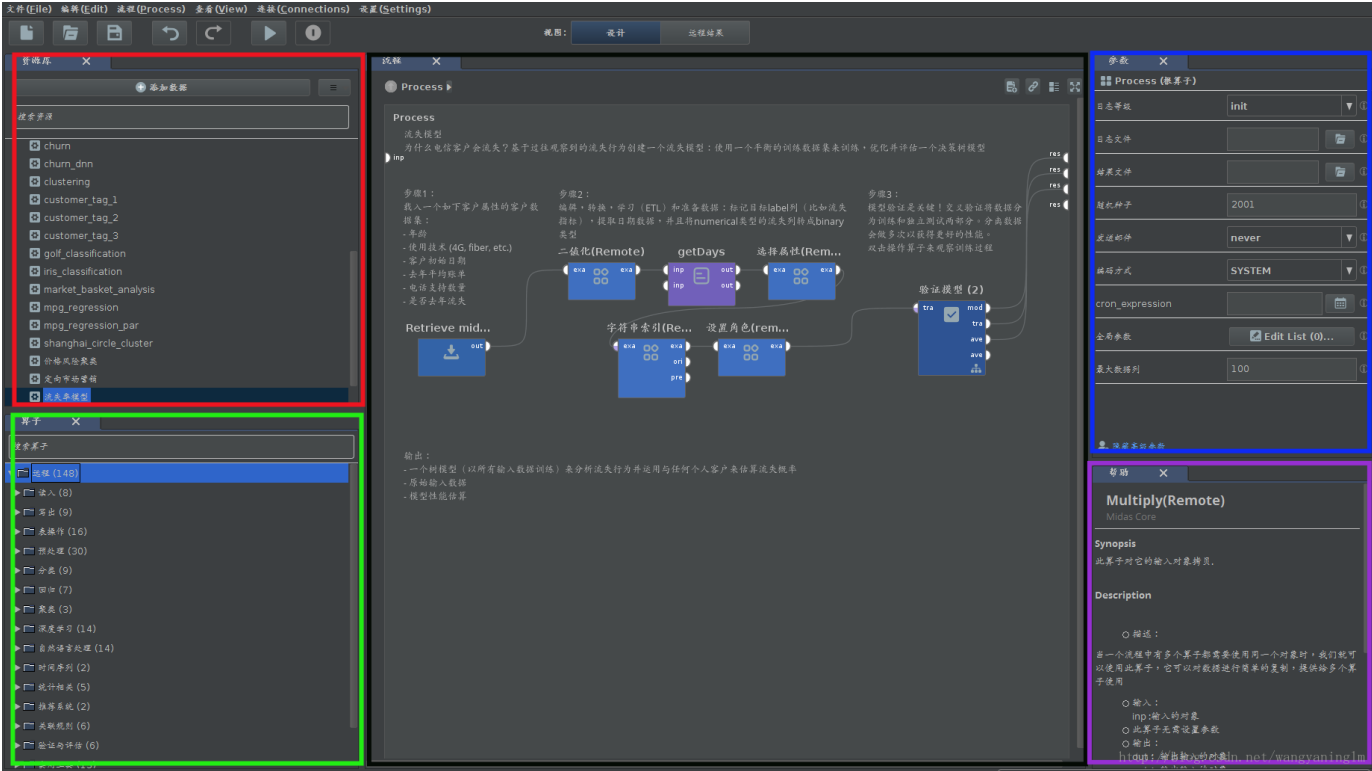


<http://blog.csdn.net/wangyaninglm>

sophon

星环还通过Transwarp Sophon来帮助数据工程师开发数据挖掘的应用。Sophon提供了可视化界面工具Midas 用于创建模型,用户只需通过拖拽数据源对象和运算符就能完成模型设计,然后将设计的模型在TDH集群上训练或预测分析。

此外,Sophon还整合了深度学习框架Tensorflow,使用户可以通过拖拽生成各种神经网络模型,灵活调参和训练,将大数据和人工智能结合起来推动业务创新。



3. 经典案例—泰坦尼克乘客的生还预测，房价预测

http://blog.csdn.net/sinat_32547403/article/details/71269804
<http://blog.csdn.net/ZengHaihong/article/details/53199559>

4. 优秀开源机器学习库

28款GitHub最流行的开源机器学习项目
<https://yq.aliyun.com/articles/30794>

15 个开源的顶级人工智能工具
<http://blog.jobbole.com/106447/>

5. 普通程序员如何学习机器学习

<https://www.zhihu.com/question/51039416>

未完待续。。。



扫码向博主提问



shiter



博客专家

天道酬勤，共同成长

擅长领域： 大数据 机器学习 计算机视觉

[去开通我的Chat快问](#)

版权声明：本文为博主原创文章，未经博主允许不得转载。 <https://blog.csdn.net/wangyaninglm/article/details/78311126>

文章标签：[计算机视觉](#) [机器学习](#) [c语言](#) [python](#)

个人分类：[机器学习](#)

相关热词：[0机器学习](#) [机器学习和](#) [机器学习的](#) [机器学习-](#) [《机器学习](#)

想对作者说点什么？

我来说两句

机器学习入门-第一个小项目

195

第一个机器学习样例该问题来自Coursera上的斯坦福大学机器学习课程，其叙述如下：现有47个房子的面积和价格，需要建立一个...

《Scikit-Learn与TensorFlow机器学习实用指南》第2章 一个完整的机器学习项目

891

第2章 一个完整的机器学习项目 来源：ApacheCN《Sklearn 与 TensorFlow 机器学习实用指南》翻译项目 译者：@SeanCheney 校...

机器学习之特征工程 - CSDN博客

5-27

做了一段机器学的东西,一直对特征这块很模糊,有时候凭感觉在找特征,然后做一些简单的变换。这篇文章的目的是系统的讲解特征的获...

机器学习中的特征工程 - CSDN博客

7-16

引言在之前学习机器学习技术中,很少关注特征工程(Feature Engineering),然而,单纯学习机器学习的算法流程,可能仍然不会使用这些算...

crm客户管理系统

百度广告



Python & 机器学习之项目实践

1.5万

机器学习是一项经验技能，经验越多越好。在项目建立的过程中，实践是掌握机器学习的最佳手段。在实践过程中，通过实际操作加...

总结:sklearn机器学习之特征工程 - CSDN博客

7-18

有这么一句话在业界广泛流传:数据和特征决定了机器学习的上限,而模型和算法只是逼近这个上限而已。那特征工程...

所以,你自认为是一个合格的机器学习工程师? - CSDN博客

6-30

在中小型企业,公司希望机器学习工程师可以独挑大梁....这个软件的透视功能上花了10年时间,常年维护几个C...

一个完整机器学习项目流程总结

2309

1. 实际问题抽象成数学问题这里的抽象成数学问题，指的我们明确我们可以获得什么样的数据，目标是一个分类还是回归或者是聚...

机器学习项目实战:泰坦尼克号获救预测

4406

```
import pandas titanic = pandas.read_csv("D:\\test\\titanic_train.csv") #进行简单的统计学分析 print titanic.des...
```

机器学习之特征工程以及特征选择的工..._CSDN博客

7-17

由此可见,特征工程尤其是特征选择在机器学习中有相当重要的地位。 通常而言,特征选择是指选择获得相应模型和算法最好性能的特...

机器学习中的「特征工程」到底是什么? - CSDN博客

5-22

1 特征工程是什么? 有这么一句话在业界广泛流传:数据和特征决定了机器学习的上限,而模型和算法只是逼近这个上限而已。那特征工...

2017年度盘点：Github上十大有趣的机器学习项目（文末有惊喜.....）

4739

微信公众号关键字全网搜索最新排名【机器学习算法】：排名第一【机器学习】：排名第一【Python】：排名第三【算法】：排名...

农村刘师傅辞去工作去做这个，半年买房买车！

新区惠众 · 顶新

机器学习在工程中的实际应用 - CSDN博客

机器学习之特征工程 - CSDN博客 原文出处:http://blog.csdn.net/dream_angel_z/article/details/49388733 版权声明:本文为博主原创文章,转载请注明来源。 目录(?)[-] 特...	5-23
28款GitHub最流行的开源机器学习项目 28款GitHub最流行的开源机器学习项目 (一) : TensorFlow排榜首 readygo 浏览 673 2016-04-18 20:31:12 机器学习开源githubT...	 1.4万
初学者必读的八个趣味机器学习项目 教程+数据集 抽时间做项目是最好的一种投资方式,在项目中你会享受学习、保持积极性并能获得更快的进展。没有任何理论可以代替实践,虽然...	 3009
机器学习的特征工程 - CSDN博客 特征工程师使用专业背景知识和技巧处理数据,使得特征能在机器学习算法上发挥更好的作用的过程。特征工程的意味着更强的灵活度,...	6-27
机器学习之sklearn特征工程 - CSDN博客 在机器学习界,有这么一种说法:数据和特征决定了机器学习的上限,而算法和模型只是逼近这个上限。所谓特征工程,是指使用专业的知...	6-7
资源 2017年GitHub中Top 30开源机器学习项目 开源机器学习项目是开放资源的一种重要类型,为用户和学习者提供了极大的便利。近期,Mybridge在对比了过去一年中机器学习...	 1653
Github上的十大机器学习项目 原文作者: Matthew Mayo 译文地址: Top 10 Machine Learning Projects on Github 文章译者: 赵屹华, 搜狗计算广告工程师, 前生...	 609
机器学习项目开发经验 机器学习项目中,可用那么多算法,那么多Feature,模型有那么多参数,怎么选择合适的?凭直觉?...	 1457
机器学习之项目实践 机器学习是一项经验技能,经验越多越好。在项目建立的过程中,实践是掌握机器学习的最佳手段。在实践过程中,通过实际操作加...	 768
机器学习项目流程清单 机器学习流程	 379
50万码农评论: 英语对于程序员有多重要! 不背单词和语法,老司机教你一个数学公式秒懂天下英语	 是的,没错
一文了解 2018年最火爆的30个机器学习项目 点击上方“CSDN”,选择“置顶公众号”关键时刻,第一时间送达!机器学习是当前最为火爆的话题之一,机器学习的开源项目也层出...	 1969
30个机器学习项目 2018年03月20日 14KB <div>下载</div>	
2017最受欢迎的十大机器学习项目 目录(?)[+]微信公众号关键字全网搜索最新排名【机器学习算法】:排名第一【机器学习】:排名第一【Python】:排名第三【算法...	 2886
【机器学习】超详细!上线一个机器学习项目你需要哪些准备? ...	 1066
20 个顶尖的 Python 机器学习开源项目 1 Scikit-learn www.github.com/scikit-learn/scikit-learn Scikit-learn 是基于Scipy为机器学习建造的的一个Python	 3123
「码字计划」: 拿万元写作基金! Python薪资多少?	python机器学习
	登录 注册

机器学习实战项目决策树完整项目

2018年03月21日 138KB 下载



机器学习项目实践流程

机器学习是一项经验技能，经验越多越好。在项目建立的过程中，实践是掌握机器学习的最佳手段。在实践过程中，通过实际操作加...

43

机器学习开源项目

最近，Mybridge发布了一篇文章，对比了过去一年中机器学习领域大约8800个开源项目后，选出30个2017年度优秀的开源项目，包...

258

对于程序员来说，英语到底多重要？

不背单词和语法，一个公式秒懂英语！



机器学习怎么入门——自学篇

时也势也，最近因为机缘巧合，下定主意要从项目管理领域跳转到机器学习领域（人工智能/深度学习），为了更快更高效的进入学...

3517

一名本科生如何入门学习机器学习

谢邀写这篇blog哈哈，一直没有写blog的想法，平时都是自己边学习边把知识点总结到纸上。关于学习方法的问题，我个人觉...

5713

如何入门Python与机器学习

编者按：本书节选自图书《Python与机器学习实战》，Python本身带有许多机器学习的第三方库，但本书在绝大多数情况下只会用...

7775

《机器学习有意思！01》- 世界上最简单的机器学习入门

本文首发于https://jizhi.im/blog/post/ml_is_fun_01 你是否也曾听人们谈起机器学习但是只有一个朦胧的概念？你是否厌倦了在同事的...

2.2万

3分钟了解入门「机器学习」该学习什么？（上）

本文来自作者 粽子 在 GitChat 上分享「零基础的新手，如何入门机器学习？」，「阅读原文」查看交流实录 「文末高能」 编辑 | 坂...

5521

订单管理系统

什么是订单管理系统



本科生如何入门学习机器学习

谢邀写这篇blog哈哈，一直没有写blog的想法，平时都是自己边学习边把知识点总结到纸上。关于学习方法的问题，我个人觉得很...

316

机器学习该如何入门

引言 可能你对这个名字叫“机器学习”的家伙不是特别的了解，但是相信用过iPhone的同学都知道iPhone的语音助手Siri，它能帮...

8273

如何学习机器学习的一点心得

结合自己的学习经历，总结一下如何学习机器学习。我自己的学习过程其实是非常混乱和痛苦的，一个人瞎搞现在也不知道入没入门...

5543

机器学习入门好文，强烈推荐

转自 飞鸟各投林 史上最强----机器学习经典总结---入门必读----心血总结-----回味无穷 让我们从机器学习谈起 导读：在本篇文章中，...

11.6万

关于机器学习，你应该至少学习这8个落地案例 | 干货集锦

1、机器学习在工程项目中的应用实践经验谈 目前，人工智能、机器学习在很多人眼中是数据科学家们的专属武器，能够应用和掌握...

1358

程序猿不会英语怎么行？英语文档都看不懂！

不背单词和语法，一个公式教你读懂天下英文→



2017.0302 算法工程师机器学习第一面

2498

投了阿里的算法工程师的实习，周一晚上打电话说约周三晚上面试，想着自己没有时间复习想往后推一推，好吧就推后了一天。 年...

机器学习算法工程师--实习面经

1727

最近找实习，整理一下面的比较完整的几家公司的的问题，如果后面还有比较完整的流程再添加吧。 小米面试 一面 给用户推荐一首歌...

一个完整的机器学习项目在Python中的演练（一）

104

磐石 大家往往会选择一本数据科学相关书籍或者完成一门在线课程来学习和掌握机器学习。但是，实际情况往往是，学完之后反而...

趣味机器学习入门小项目（附教程与数据）

652

微信公众号 关键字全网搜索最新排名 【机器学习算法】：排名第一 【机器学习】：排名第一 【Python】：排名第三 【算法】：排...

每周一个机器学习小项目001：全链接层实现与分类问题

62

打算更新一个小的系列，不定期更新，但是题目定为《每周一个机器学习小项目》，主要偏重于深度学习内容，穿插一些传统的机器...

没有更多推荐了，返回首页

个人资料



shiter



博客专家

关注

向TA提问

原创

198

粉丝

2535

喜欢

526

评论

809

等级： 博客 7

访问： 155万+

积分： 1万+

排名： 953

勋章：

儿童编程



公众号，github

公众号： 老王和他的IT界朋友们

欢迎投稿： shiter@live.cn

QQ交流群： 593683975

加群问题：抛硬币正面上的期望？

我们想用一段音乐，几张图片，
些许文字绘制的IT圈心路历程，
为攻城狮，为程序员带来更多的人文关怀。

登录

注册

×

「码字计划」：拿万元写作基金！ Python薪资多少？

python机器学习



github: <https://github.com/wynshiter>

码云地址: <https://gitee.com/wangyaning/>

最新文章

- 自然语言处理简介（1） ---- 服务梳理与传统汉语分词
- 浅谈pandas , pyspark 的大数据ETL实践经验
- 设限
- 你是不是傻，怎么不选最好的？
- 均值与期望：傻傻分不清？

博主专栏

-  计算机视觉OpenCV实现
阅读量：537185 46 篇
 -  跨平台通信系统
阅读量：26282 6 篇
 -  图论及其算法在计算机视觉中的应用
阅读量：85076 10 篇
 -  OpenCV特征点检测
阅读量：91743 7 篇
- leetcode c++&&pytho
- [展开](#)

个人分类

- 图形图像 61篇
 - OpenGL 8篇
 - OpenCV 63篇
 - Linux功能实现 9篇
 - windows功能实现 24篇
- [展开](#)

归档

- 2018年7月 1篇
- 2018年6月 1篇
- 2018年5月 4篇
- 2018年4月 2篇
- 2018年3月 1篇