

[学习](#) [实践](#) [活动](#) [专区](#) [工具](#) [TVP](#) [文章/答案/技术大牛](#)[写文章](#)[提](#)[专栏首页](#) [量化小白上分记](#) [符号回归和遗传规划](#)

符号回归和遗传规划

发布于2019-10-17 17:32:45 阅读 3.7K

背景

回归分析是一种常用的统计方法，用来分析自变量和因变量的线性相关关系，在线性回归分析中，变量间的关系形式是确定的，只需要对关系式的系数做出估计。

符号回归是另一种回归分析方法，不同之处在于，符号回归不对变量间的关系形式做出假设，因此，符号回归过程既包括函数形式/运算符的确定，也包括函数中变量系数的确定。符号回归相比于线性回归的优势在于可以挖掘变量间可能存在的非线性关系，但也可能过拟合。

遗传规划（Genetic Program）是一种求解符号回归的方法。本文使用python的gplearn包实现符号回归。

数据准备

本文内容主要参考[1]。假设现在有两个自变量 X_1, X_2 ，与因变量的关系为

$$y = X_0^2 - X_1^2 + X_1 - 1$$

随机生成100个 X_0, X_1, y 的随机变量，50个做trainset，50个做testset，用遗传规划的方式来推测变量之间的关系式，作为对比，也用随机森林和决策树的算法做一遍看效果。

上述非线性的方程在 X_0, X_1 取值 $[-1, 1]$ 的情况下，图像如下，代码文件获取后台回复“gplearn”。

```
1 from gplearn.genetic import SymbolicRegressor
2 from sklearn.ensemble import RandomForestRegressor
3 from sklearn.tree import DecisionTreeRegressor
4 from sklearn.utils.random import check_random_state
5 from mpl_toolkits.mplot3d import Axes3D
6 import matplotlib.pyplot as plt
7 import numpy as np
8 import graphviz
9 import pydotplus
10
11 # plot
12 x0 = np.arange(-1, 1, 1/10.)
13 x1 = np.arange(-1, 1, 1/10.)
14 x0, x1 = np.meshgrid(x0, x1)
15 y_truth = x0**2 - x1**2 + x1 - 1
16
17 ax = plt.figure(figsize = (10,7)).gca(projection='3d')
18 ax.set_xlim(-1, 1)
19 ax.set_ylim(-1, 1)
20 surf = ax.plot_surface(x0, x1, y_truth, rstride=1, cstride=1,
21                       color='green', alpha=0.5)
22 plt.show()
```

背景

数据准备

[符号回归和遗传规划](#)[作者介绍](#)[参考文献](#)

量化小白

关注

文章	阅读量	获赞
79	172.2K	266

精选专题

腾讯云原生专题
云原生技术干货地。

活动推荐

社区创作者年终回顾
参与活动，赢取限量周边礼品

邀请好友加入自媒体分...
邀请好友，同享奖励 30 / 100 / 180 元云服务...

运营活动



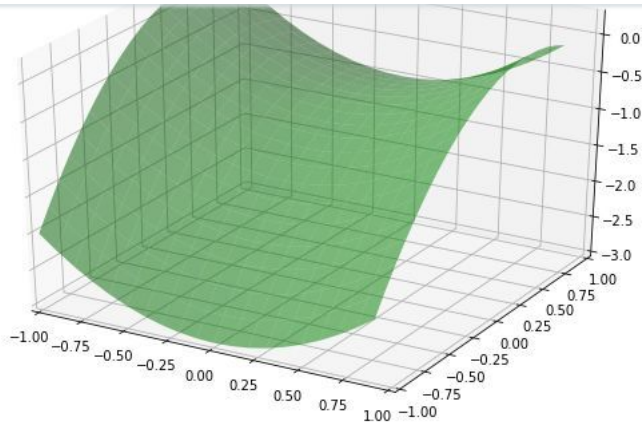
关注

腾讯云开发者公众号

将获得

10元无门槛代金券
洞察腾讯核心技术
剖析业界实践案例

- 背景
- 数据准备
- [gplearn 遗传规划](#)
- 参考文献



在曲面上进行随机数模拟生成训练集和测试集

```

1 | rng = check_random_state(0)
2 |
3 | # Training samples
4 | X_train = rng.uniform(-1, 1, 100).reshape(50, 2)
5 | y_train = X_train[:, 0]**2 - X_train[:, 1]**2 + X_train[:, 1] - 1
6 |
7 | # Testing samples
8 | X_test = rng.uniform(-1, 1, 100).reshape(50, 2)
9 | y_test = X_test[:, 0]**2 - X_test[:, 1]**2 + X_test[:, 1] - 1

```

gplearn 遗传规划

关于gplearn就不过多说明了，个人感觉它是python中快速实现遗传规划最好用的包，类似的包还有deap，但deap如果要向量的模型，需要改很多代码，放弃了。

用遗传规划的方式实现符号回归总体来说是一个优化问题，首先，最优化的目标是模型预测值和真实值之间的MSE或MAE或RMSE，跟机器学习算法是一样的。

其次，符号回归需要事先给出一些运算符号，gplearn中内置的运算符包括加减乘除、开方、对数、三角函数等，详见下图，用户也可以自定义一些运算符

量化小白
79 篇文章

符号回归和遗传规划

include strings to indicate either individual functions as outlined below, or you can also include your own functions as built using the `make_function` factory from the `functions` module.

Available individual functions are:

- 'add' : addition, arity=2.
- 'sub' : subtraction, arity=2.
- 'mul' : multiplication, arity=2.
- 'div' : protected division where a denominator near-zero returns 1., arity=2.
- 'sqrt' : protected square root where the absolute value of the argument is used, arity=1.
- 'log' : protected log where the absolute value of the argument is used and a near-zero argument returns 0., arity=1.
- 'abs' : absolute value, arity=1.
- 'neg' : negative, arity=1.
- 'inv' : protected inverse where a near-zero argument returns 0., arity=1.

[展开阅读全文](#)

Python

编程算法

线性回归

决策树

举报

点赞 3

分享

[登录](#) 后参与评论

1 条评论

[最新](#) [高赞](#)



用户9812490

2022-06-08

请问从哪里获得代码呢?

0 回复

相关文章

普林、DeepMind新研究：结合深度学习和符号回归，从深度模型中看...

来自普林斯顿、DeepMind 等机构的研究人员提出了一种解决方案：结合深度学习和符号回归实现这一目标。



机器之心

目录

背景

数据准备

[gplearn](#) [遗传规划](#)

参考文献



量化小白
79 篇文章

符号回归和遗传规划

其中 c 和 x 为 n 维列向量, A 、 Aeq 为适当维数的矩阵, b 、 beq 为适当维数的列向量。

全栈程序员站长

Math-Model (一) 算法综述

美赛马上涨了, 总结一下这些年参赛的算法(我打编程位), 数学建模主要模型不单独写, 参考数学模型第四版教材即可, 只给出编程中一些重要的算法目录, 如果有方法漏写, 请评...

Pulsar-V

回归分析初步 (一)

回归: 回归 (Regression) 过程主要研究和解决的问题是识别和分析出隐藏在实验数据中的数学关系。回归作为科学研究的一种手段, 通过获取已有的实验数据, 获取某个...

马远良

特征工程系列: 笛卡尔乘积特征构造以及遗传编程特征构造

关于作者: JunLiang, 一个热爱挖掘的数据从业者, 勤学好问、动手达人, 期待与大家一起交流探讨机器学习相关内容~

木东居士

基于遗传规划算法的CODLAG推进系统燃气轮机轴系扭矩和燃油流量估算 (AI)

本文利用可公开获得的基于状态的舰船柴油-燃气和燃气-天然气联合推进系统 (CODLAG) 推进系统的数据集, 以获取符号表达式, 从而可以使用遗传编程 (GP) 算法估算燃...

田冠宇

数学建模的一些方法_对数学建模的认识

本人在大学时期 待了两年的数学建模社团, 也参加过国赛, 最近有些许感性, 想以此纪念一下。

全栈程序员站长

回归分析 (二) ——符号回归

符号回归 (Symbolic Regression) 作为一种监督学习方法, 试图发现某种隐藏的数学公式, 以此利用特征变量预测目标变量。符号回归的优点就是可以不用...

马远良

孟德尔随机化之Wald ratio方法 (二)

有时候, IV可能无法将遗传亚组依据不同暴露层次单纯分为两组。这时候, 一般会出现3个遗传亚组: AA (主要纯合子,majorhomozygote), Aa (杂合子,h...

生信与临床

遗传算法系列之二:“欺骗”深度学习的遗传算法

这篇博客主要介绍不同问题的遗传算法。遗传算法是通用的全局优化算法, 因此有很多的应用。有很多应用我是看不懂的, 比如机器人步态优化。机器人...

AlgorithmDog

方程就是二叉树森林? 遗传算法从数据中直接发现未知控制方程和物...

机器之心专栏 机器之心编辑部 偏微分方程是领域知识的一种简洁且易于理解的表示形式, 对于加深人类对物理世界的认知以及预测未来变化至关重要。然而, 现实世界的系统过于...

目录

背景

数据准备

[gplearn 遗传规划](#)

参考文献



量化小白
79 篇文章

符号回归和遗传规划

目录

- 背景
- 数据准备
- [gplearn 遗传规划](#)
- 参考文献

【机器学习】机器学习背后的统计思想

要理解机器学习首先要明白他的定义，很多人给出了定义“学习是人类具有的一种重要智能行为，但究竟什么是学习，长期以来却众说纷纭。社会学家、逻辑学家和心理学家都各有其...

陆勤_数据人网

Math-Model算法综述

数学建模主要模型不单独写，参考数学模型第四版教材即可，只给出编程中一些重要的算法目录，如果有方法漏写，请评论区指出，笔者添加，谢谢QAQ

Pulsar-V

遗传进展系列 | 1. 回归系数 相关系数 遗传力的关系

回归系数: 两个变量的关系, 构建回归模型, $y = a + b*x$, b为回归系数

邓飞

学界 | 循环神经网络自动生成程序：谷歌大脑提出「优先级队列训练」

机器之心

实际工作中，算法工程师需要哪些技能？

对机器学习感兴趣是一回事，实际上开始在现场工作是另一回事，实际中，真正开始从事机器学习工程师工作的整体思维方式和具体技能，又是怎样的呢？

double

【Python环境】python中的数据分析和机器学习库

Python 计算机视觉 SimpleCV—开源的计算机视觉框架，可以访问如OpenCV等高性能计算机视觉库。使用Python编写，可以在Mac、Windows...

陆勤_数据人网

【统计学家的故事】现代统计科学的创立者：卡尔·皮尔逊

卡尔·皮尔逊 (Karl Pearson, 1857年3月27日~1936年4月27日) 是英国数学家，生物统计学家，数理统计学的创立者，自由思想者，对生物统计学、气...

统计学家

[置顶] 机器学习简介及常用算法

概念 什么是机器学习？机器学习是英文名称Machine Learning(简称ML)的直译。机器学习涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多...

xiangzhihong

[更多文章](#)



量化小白
79 篇文章

符号回归和遗传规划

目录

专栏文章	自媒体分享计划	技术周刊	视频介绍	背景 数据准备 gplearn 遗传规划 参考文献
阅读清单	邀请作者入驻	社区标签	社区规范	
互动问答	自荐上首页	开发者手册	免责声明	
技术沙龙	技术竞赛	开发者实验室	联系我们	
团队主页			友情链接	

热门产品	域名注册	云服务器	区块链服务	消息队列	网络加速	云数据库	域名解析	云存储
热门推荐	人脸识别	腾讯会议	企业云	CDN 加速	视频通话	图像分析	MySQL 数据库	SSL 证书
更多推荐	数据安全	负载均衡	短信	文字识别	云点播	商标注册	小程序开发	网站监控

扫
领