

机器学习

- [目录视图](#)
- [摘要视图](#)
- [订阅](#)
- [发布Chat](#)

[异步赠书：10月Python畅销书升级](#) [【线路图】人工智能到底学什么？！](#) [程序员10月书讯](#) [每周荐书（京东篇）：618取胜之道、质量保障、技术解密](#)

LabelEncoder OneHotEncoder 在python中的运用

2016-12-12 11:19 1228人阅读 [评论\(0\)](#) [收藏](#) [举报](#)

版权声明：本文为博主原创文章，未经博主允许不得转载。

LabelEncoder 是对不连续的数字或者文本进行编号
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
le.fit([1,5,67,100])
le.transform([1,1,100,67,5])
输出：array([0,0,3,2,1])

OneHotEncoder 用于将表示分类的数据扩维：
from sklearn.preprocessing import OneHotEncoder
ohe = OneHotEncoder()
ohe.fit([[1],[2],[3],[4]])
ohe.transform([2],[3],[1],[4]).toarray()
输出：[[0,1,0,0] , [0,0,1,0] , [1,0,0,0] , [0,0,0,1]]

顶

0

踩

0

-
-

关闭

- [上一篇python实现的四种抽样方法](#)
- [下一篇python pandas中的random模块一些函数的用法](#)

相关文章推荐

- [CTR——人工神经网络+决策树](#)
- [深度学习部署系统构建--刘文志](#)
- [机器学习 数据预处理之独热编码 \(One-Hot Encoding\)](#)
- [搜狗机器翻译技术分享--陈伟](#)
- [基于sklearn 的one hot encoding](#)
- [Hadoop生态系统零基础入门](#)
- [基于【pandas】的数据预处理（含定制【OneHotEncoder】方法）](#)
- [最懂程序员的学习方式 TensorFlow入门](#)

- [one hot coding -机器学习](#)
- [Retrofit 从入门封装到源码解析](#)

查看评论
处理离散型特征和连续型特征共存的情况 归一化 论述了对离散特征进行one-hot编码的意义

* 以上用户言论只代表其个人钱柜娱乐开户，不代表CSDN网站的钱柜娱乐开户或立场

- [程序员如何转型AI工程师--蒋涛](#)
- [数据处理系列：（六）标签特征二元化](#)
- [机器学习实验（十三）：90%的准确率，根据中文名字预测性别！](#)
- [数据处理——One-Hot Encoding](#)
- [wang1127248268的数据预处理（含定制【OneHotEncoder】方法）](#)

- 访问：38891次
- 积分：985
- 等级：
- 排名：千里之外
- 原创：35篇
- 转载：107篇
- 译文：9篇
- 评论：4条

文章搜索

搜索

文章分类

关闭

- [opencv](#)(0)
- [大数据](#)(3)
- [数据库性能优化](#)(2)
- [数据结构](#)(4)
- [hbase](#)(1)
- [设计模式](#)(1)
- [面试题](#)(1)
- [sklearn](#)(1)
- [spark](#)(1)
- [机器学习](#)(28)
- [特征选择](#)(2)
- [数据挖掘过程](#)(1)
- [ROC曲线](#)(1)
- [机器学习中的距离](#)(1)
- [进程通信方式](#)(1)
- [hbase用法](#)(1)
- [redis的用法](#)(1)
- [mongodb和mysql的比较](#)(1)
- [mongodb](#)(1)
- [循环神经网络](#)(1)
- [nlp](#)(2)
- [神经网络-深度学习](#)(1)
- [算法优化](#)(1)
- [文本挖掘](#)(19)
- [推荐系统](#)(3)
- [python](#)(1)
- [深度学习](#)(24)
- [数据挖掘](#)(1)
- [神](#)(0)
- [java基础](#)(1)
- [用户画像](#)(1)
- [hadoop](#)(1)
- [li](#)(0)
- [领悟](#)(1)
- [几区xue](#)(0)
- [maven](#)(1)
- [数据库](#)(0)
- [知识图谱](#)(1)
- [java](#)(4)

[关闭](#)

- [舆情分析](#)(2)
- [图数据库](#)(1)
- [最小编辑距离](#)(1)
- [elasticSearch 文档](#)(1)
- [Dao](#)(1)

文章存档

- [2017年10月](#)(5)
- [2017年09月](#)(36)
- [2017年08月](#)(78)
- [2017年01月](#)(1)
- [2016年12月](#)(9)
- [2016年11月](#)(22)

阅读排行

- [python实现的四种抽样方法](#)(5147)
- [自动化提取关键字](#)(3710)
- [sklearn的一些总结](#)(3579)
- [舆情分析的几个概念](#)(2309)
- [数据挖掘中的支持度 置信度 期望置信度 提升度](#)(1733)
- [sql高级函数汇总](#)(1441)
- [特征选择方法](#)(1301)
- [LabelEncoder OneHotEncoder 在python中的运用](#)(1225)
- [teradata sql学习笔记](#)(909)
- [XGBOOST参数说明](#)(897)

评论排行

- [sklearn的一些总结](#)(1)
- [数据挖掘中的支持度 置信度 期望置信度 提升度](#)(1)
- [sklearn的一些技巧总结](#)(1)
- [Keras搭建的自编码模型](#)(1)
- [spark-streaming学习笔记总结](#)(0)
- [elasticsearch的API](#)(0)
- [spark在yarn上面的运行模型：yarn-cluster和yarn-client两种运行模式](#)(0)
- [hbase的基本操作](#)(0)
- [最小生成树-Prim算法和Kruskal算法](#)(0)
- [hive调优总结](#)(0)

关闭

推荐文章

- [* CSDN邀请您来GitChat赚钱啦！](#)
- [* 行为驱动开发（BDD）你准备好了吗？](#)
- [* 如何更加安全、高效地利用开源项目？](#)
- [* 程序员业余时间修炼指南](#)
- [* DevOps 在公司项目中的实践落地](#)
- [*Jenkins + Django 完整实战，细化到每一步操作](#)

最新评论

- [Keras搭建的自编码模型](#)
[qq_37078891](#): 请问如何在自编码后面添加一个分类器呢？谢谢
- [数据挖掘中的支持度 置信度 期望置信度 提升度](#)
[qq_20276233](#): 你这个背景“真好看”啊！！
- [sklearn的一些技巧总结](#)
[YiRanZhiLiPoSui](#): 转载不知道注明来源？
- [sklearn的一些总结](#)
[Alis_xt](#): /alis_xt/article/details/61419...

关闭

关闭