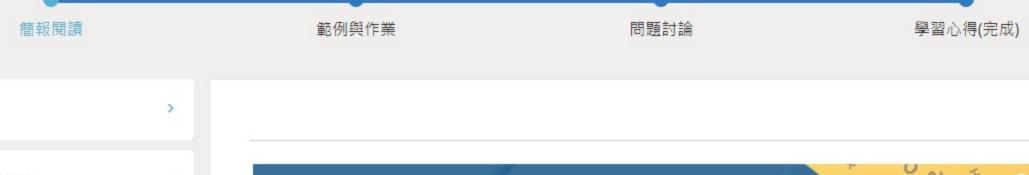
അവ AI共學社群 我的 AI共學社群 > Python資料科學 > D09 使用 Pandas 讀寫各種常用的檔案格式

D09 使用 Pandas 讀寫各種常用的檔案格式

囯







E



5.0 3.6

0.2

0.2

0.4

0

0.2 1 1.4 2 1.3 0.2 3 1.5 0.2

sepal length sepal width petal length petal width target

	featrue1	featrue2	featrue3	featrue4	target
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
5	5.4	3.9	1.7	0.4	0

data

0

3

4.9 3.0 1.4 1 2 4.7 3.2 1.3

ZN INDUS CHAS NOX

17.8 392.83

18.7 394.63

18.7 396.90

2 242

3 222

4 222

2.31

7.07

7.07

2.18

5.1

4.6

5.0

pd.read_excel() 只會讀第一個工作簿,如今天 excel 工作簿不只一個,我們可以使用 sheetname 參

RM AGE

0 0.469 6.421 78.9 4.9671

0 0.469 7.185 61.1 4.9671

65.2 4.0900

45.8 6.0622

0 0.538 6.575

0 0.458 6.998

4.03

2.94

5.33

0 0.458 7.147 54.2 6.0622 2.18 3 222 18.7 396.90 ,usecols=['TAX','PTRATIO','B','LSTAT']) TAX PTRATIO B LSTAT 15.3 396.90 4.98 17.8 396.90 9.14

TAX PTRATIO B LSTAT 0 296 15.3 396.90 4.98 1 242 17.8 396.90 9.14

17.8 392.83

18.7 394.63

18.7 396.90

boston_data.to_json('boston.json')

connection = sqlite3.connect('sql db.sqlite') boston_data.to_sql('boston', connection, if_exists='replace') connection.commit() connection.close() 讀取 SQLite3 資料庫可以使用 pd.io.sql.read_sql,可以直接下 SQL 指令對 sql_db 中的 boston 做搜 connection = sqlite3.connect('sql_db.sqlite')

B LSTAT

4.98

9.14

4.03

2.94

15.3 396.90

17.8 396.90

17.8 392.83

18.7 394.63

python pandas 讀取檔案——讀取具有明顯分隔符的

▶ 資料載入的主要來源 * 文字檔 * 資料庫 * 炯頁 (1) 文字格式資料 (Text format) * pandas 提供讀取檔案並存為 DataFrame 的函式: 語法:pd.<func_name>(<poram>):例如pd.read_csv('<file_path>') 前式 說明 read_cav() 從檔案或 URL 讓取預設以短點分隔的資料 read_table() 使機能或 URL 讓取預設以定位總 ("t") 點分隔的資料 read_fw(t) 讓取開定度度機位格式 (Fixed-width column format,沒有分隔符號) read_excel() 從 XLS 或 XLSX 檔讀取表格資料 read_hdf() 譲取 HDF5 植
read_html() 譲取 HTML 檔案裡的所有表格資料 read_json() 譲取 JSON 檔 read_msgpack() 顕取 MessagePack 二元格式檔 read_pickle() 譲取以 pickle 格式儲存的 Python 物件 護取 SAS 轄 譲取 SQL 查詢結果 read_sas() read_sql() 延伸閱讀

3 ID000004E Male Mumbai 35000 4 ID000007b Male 5 ID000008I: Male Saharsa Bengaluru 100000 5 ID000009J-Male 7 ID000010k Male 9 ID000012h Male Bengaluru 20000 10 ID000013ħ Male Kochi



0.2 0 5.1 3.5 1.4 4.9 3.0 1.4 0.2 2 4.7 3.2 0.2 1.3 3 4.6 0.2 3.1 1.5 0.2

可以再資料圖取 過檔案裡放置行			旨定行的名稱	,因為有指	定行名稱	所以必須以	以 heardei	r=0 參數跳
iris_data = pd. iris_data	.read_csv('ir	ris.csv',he	eader=0,names=['featrue1',	'featrue2'	,'featrue3'	,'featrue4	','target'])

OpenPyXL 套件,所以使用前確認至少安裝其中一項。 data = pd.read_excel('data.xls')

sepal length sepal width petal length petal width target

1.4

1.5

1.4

0.2

0.2

0.2

0.2

0.2

DIS RAD TAX PTRATIO

1 296

2 242

2 242

222

3

0

0

0

0

B LSTAT

4.98

9.14

4.03

2.94

5.33

15.3 396.90

17.8 396.90

17.8 392.83

18.7 394.63

3.5

3.1

3.6

iris_data.to_csv('my_iris.csv')

Pandas 利用 pd.read_excel() 函數支援讀取 Excel 2003 之後的格式資料,此方法內部是使用 XLRD 或

pd.read_excel 和 pd.read_csv 一樣有 usecols、header、names 可以使用 boston_data = pd.read_excel('data.xls', sheet_name='boston', header=0 \

利用 pd.read_json() 函數支援讀取 JavaScript 物件表示法 (JSON) 格式資料,json 具有跨平台與程式語 boston_data = pd.read_json('boston.json') boston data

4.03

2.94

5.33

boston_data.to_excel('my_boston.xlsx',sheet_name='boston')

boston_data = pd.read_excel('data.xls', sheet_name='boston', header=0 \ ,usecols=['TAX','PTRATIO','B','LSTAT'])

python pandas 讀取檔案 ITREAD 01

I. Python 1. Python 簡介 2. 資料與模組 3. 資料結構

使用python讀取資料科學最常用的檔案格式 其他·發表 2019-02-17

12 ID000016C Male Mumbai 13 ID00001BS Female Surat 25000 14 ID000019T Female Pune 24000 15 ID000021\ Male Bhubanes 27000 16 ID000022V Female Howrah 28000

> 1引言 數據分析、數據挖掘、可視化是Python的眾多強項之一,但無論是這幾項中的 哪一項都必須以數據作為基礎,數據通常都存儲在外部文件中,例如txt、 csv、excel、資料庫。本篇中,我們來捋一捋Python中那些外部數據文件讀 取、寫入的常用方法。

Python 資料科學程式馬拉松

DataFrame 的資料型態。有時候資料太多不想一次讀取這麼多行可以使用 usercols 參數指定讀取的行 iris_data = pd.read_csv('iris.csv') iris_data

有時候資料太多不想一次讀取這麼多行可以使用 usercols 參數指定讀取的行名稱。 iris_data = pd.read_csv('iris.csv',usecols=['petal length','petal width','target']) iris_data

4 5 可以再 過檔案 iris

數傳入要讀取頁面的名稱。 boston_data = pd.read_excel('data.xls', sheet_name='boston')

boston_data

CRIM

0 0.00632 18.0

4 0.06905 0.0

0.0

0.0

0.0

1 0.02731

2 0.02729

3 0.03237

輸出 excel 檔案使用 .to_excel() 的方法就可以將資料存到指定路徑下,也可以使用 sheet_name 更改 工作簿名稱。 讀寫 json 言的可攜性。

輸出 json 檔案使用 .to_json() 的方法就可以將資料存到指定路徑下。 讀寫 SQL 資料庫

boston_data_sql = pd.io.sql.read_sql("select * from boston",connection) connection.close() boston_data_sql

index TAX PTRATIO

0 296

1 242

2 242

3 222

尋。

補充教材 網站: itread01

網站: yltang

III. 機器學習 12. 機器學習能介 13. 贩售式學習 14. 非监督式學習

> 1、什麼是檔案格式? 1 ID Gender City Monthly_1 2 ID000002C Female Delhi 20000

資料科學與機器學習

第9章 資料載入、儲存、與檔案格式

<?xml version="1.0"?>

<name>Ankit</name>

<company>Anlytics Vidhya</company>

<phone>+9187654321</phone>

"id":"2", "Name", "Esizo"

undefined

</contact-info>

重要知識點

CSV 的資料由很多文本資料組成,資料之間以逗點隔開。 首先我們使用 Pandas 的套件中 pd.read_csv 讀取一個 CSV 檔案夾中的 iris.csv,讀取之後即為 名稱。

petal length petal width target 0 1.4 1.4 1.7

輸出

讀寫 excel

3 222

4 222

任何 SQL 資料庫如果支援遵守 Python DB-API 都可以被 Pandas 讀取。以下先用 boston.csv 的資料 寫入 SQLite3 資料庫中並且命名 boston,由 if_exists 參數判斷是否存在資料庫,如果檔案不存在會立 即被建立,如果存在 if_exists='replace' 將會取代掉原本資料,if_exists='append' 將會繼續寫在原有 資料下。 import sqlite3

3 知識點回顧 • 資料型態多元,我們可以使用 Pandas 針對不同資料型態讀取,讀寫 csv(read_csv、to_csv)、 讀寫 excel(read_excel、to_excel)、讀寫 json(read_json、to_json)、讀寫 SQL 資料庫 (io.sql.read_sql、to_sql), 讀取進來後都是DataFrame的型態,方便之後進行資料的分析,也 可以藉由 Pandas 做資料型態的轉換。

資料

其他

4. 決策图图函式 5. 物件與规则 II. 资料科學 6. 資料科學能介 7. NumPy 8. pandas 9. 资料载入 10. 资料清理 11. 阅表

其他

Author : kevinelstri DateTime: 2017/3/13 檔案格式是在檔案中儲存資訊的一種標準方法。首先,檔案格式指定檔案是一個二進位制或ASCII檔案。其 次,檔案展示了檔案的組織形式。例如,逗號分隔值(CSV)檔案格式儲存在純文字的表格資料。 CSV

"Employee": ["id":"1", "Name": "Ankit", "Sal": "1000",

JSON

其他·發表 2018-12-18 關於這篇博文提到的所有函式都可以在 IO Tools (Text, CSV, HDF5, ...) 2 找到。 首先我們來列舉一下 pandas 處理檔案的函式1: ■ pd.read_csv(filepath): 從 CSV 檔案匯入資料 ■ pd.read_table(filepath): 從限定分隔符的文字檔案匯入資料 ■ pd.read_excel(filepath): 從 Excel 檔案匯入資料 ■ pd.read_sql(query, connection_object): 從 SQL 表 / 庫匯入資料 pd.read_json(json_string): 從 JSON 格式的字串匯入資料 ■ pd.read_html(url):解析 URL、字串或者 HTML 檔案,抽取其中的 tables 表格 ■ pd.read_clipboard():從你的貼上板獲取內容,並傳給 read_table() ■ pd.DataFrame(dict): 從字典物件匯入資料, Key 是列名, Value 是資料 想要讀取具有明顯分隔符的資料,我們可以採用的函式是: pd.read_csv(filepath) pd.read_table(filepath) 與此問有關的相關引數有三個: sep, delimiter, delim_whitespace, 具體含義以 pandas.read_table 為例,

使用 python 讀取資料科學最常用的檔案格式 網站: itread01 ITREAD 01

11 ID000014C Female Mumbai

下一步:閱讀範例與完成作業