

D6：EDA：欄位的資料類型介紹及處理



簡報閱讀



範例與作業



問題討論

EDA資料類型介紹 >

知識地圖 >

本日知識點目標 >

資料類型 >

重要知識點複習 >

延伸閱讀 >

推薦延伸閱讀 >

解題時間 >

EDA資料類型介紹



知識地圖

知識地圖 機器學習前處理 欄位的資料類型介紹及處理



機器學習概論 Introduction of Machine Learning

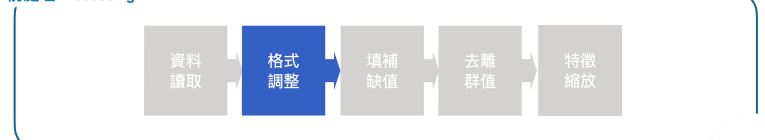
監督式學習 Supervised Learning



非監督式學習 Unsupervised Learning



前處理 Processing



本日知識點目標

- 了解 pandas dataframe 欄位的基本資料類型

資料類型

資料的欄位變數一般可分

01

- 離散變數: 只能用整數單位計算的變數
 - 房子的房間數量、性別、國家
- 連續變數: 在一定區間內可以任意取值的變數
 - Ex: 測量的身高、飛機起飛到降落所花費的時間、車速

Pandas DataFrame中 最常見的欄位資料類型有三種

02

- float64 : 浮點數，可表示離散或連續變數
- int64 : 整數，可表示離散或連續變數
- object : 包含字串，用於表示類別型變數

當然還有日期、boolean 等等不同的格式，實務中遇到再 google 就好

03

重要知識點複習

- 拿到資料的第一步，通常就是看我們有什麼，觀察有什麼欄位，這些欄位代表什麼意義、以什麼樣的資料類型來儲存
- 資料原來是字串/類別的話，如果要做進一步的分析時（如訓練模型），一般需要轉為

料的不同類別是有序的，例如該資料是年齡分組，類別有小孩、年輕人、老人，表示為 0, 1, 2 是合理的，因為年齡上老人 > 年輕人、年輕人 > 小孩

- One Hot encoding：使用時機通常是該資料的不同類別是無序的，例如國家

延伸閱讀



除了每日知識點的基礎之外，推薦的延伸閱讀能補足學員們對該知識點的了解程度，建議您解完每日題目後，若有多餘時間，可再補充延伸閱讀文章內容。

推薦延伸閱讀

Pandas 數據類型 csdn - Claroja

Pandas dtypes(數據類型)

原創 Claroja 2017-05-23 16:55:29 61245 收藏 14
分類專欄: pandas

<http://pandas.pydata.org/pandas-docs/stable/basics.html?highlight=astype#selecting-columns-based-on-dtype>

Pandas所支持的數據類型:

1. float
2. int
3. bool
4. datetime64[ns]
5. datetime64[ns, tz]
6. timedelta[ns]
7. category
8. object

默認的數據類型是int64,float64.

在後面的課程會再提及。

Pandas 所支持的數據類型

1. float
 2. int
 3. bool
 4. datetime64 [ns]
 5. datetime64 [ns.tz]
 6. timedelta [ns]
 7. category
 8. object
- 默認的數據類型是 int64,float64

解題時間



Sample Code & 作業
開始解題



[下一步：閱讀範例與完成作業](#)