

D10：數值型特徵 - 去除離群值



Day 10 機器學習前處理
數值型特徵-去除離群值

CUPOY

陳明佑
出題教練

知識地圖 數值型特徵 - 去除離群值

CUPOY

機器學習前處理

監督式學習
Supervised Learning

前處理
Processing

探索式
數據分析
Exploratory
Data
Analysis

特徵
工程
Feature
Engineering

模型
選擇
Model
selection

參數調整
Fine-tuning

集成
Ensemble

非監督式學習
Unsupervised Learning

分群
Clustering

降維
Dimension
Reduction

[PDF 下載](#)

[全螢幕](#)

Sample Code & 作業內容

請參閱範例Day_010_Outliers.ipynb：房價預測

作業1：試著限制'1樓地板面積(平方英尺)' (1stFlrSF) 欄位的上下限，看看能否再進一步提高分數？

作業2：續前題，去除離群值有兩類方式：捨棄離群值(刪除離群的資料) 以及調整離群值，請試著用同樣的上下限，改為'捨棄離群值'的方法，看看結果會變好還是變差？並試著解釋原因。

作業請繳交Day_010_HW.ipynb

[檢視範例](#)

提交作業



[如何提交](#) ▾

熱門問答

[William Mok](#) · 2019.09.14 · 2 回答

[作業 1,2 出 Error](#)

作業 1,2 出了以下 error: C:\Users\user\Anaconda3\lib\site-packages\scipy\stats\stats.py:1713: FutureWarning: Using a non-tuple sequence for multidimensional indexing is deprecated; use...

[黃世安](#) · 2019.09.10 · 1 回答

[Day10的範例](#)

[99請問我只是照著執行一遍最後的圖就出現這個問題了](#)

[陳政陽](#) · 2019.09.10 · 1 回答

[D10: 試著將 max 取代為 q99](#)

`app_train[app_train['AMT_ANNUITY'] == app_train['AMT_ANNUITY'].max()] = np.percentile(app_train[~app_train['AMT_ANNUITY'].isnull()]['AMT_ANNUITY'], q = 99)`在此段程式碼中,...

[查看更多](#)

到 Cupoy 問答社區提問，讓教練群回答你的疑難雜症

[向專家提問](#)[如何提問](#) ▾