

## 建立迴歸模型以預測賽爾提克季後賽的得分

數據探索  
可視化

1. 得分的統計量與盒鬚圖
2. 得分是否呈現常態?(常態檢定)
3. 透過相關係數探討得分與那些變數有高度相關性( $r > 0.5$ )
4. 得分與非連續型變數是否有相關?(對戰球隊、主客場、年度……)

Data rolling  
and shift

由於發現得分(PTS)與, 投籃命中率(FG\_PCT), 三分球進球數(FG3M), 助攻(AST), 正負分值(PLUS\_MINUS)有高度相關。

故增加前五場平均\_\_\_\_\_ 與前一場\_\_\_\_\_以及前五場對戰此對手平均\_\_\_\_\_ 與前一場\_\_\_\_\_。

#增加欄位後記得檢查遺失值

特徵重要性排序

將所有 MA\_5 與 LAST\_5 做連續型特徵篩選, 得出 True 的變數, 並透過圖表畫出由高到低的特徵重要性。

特徵工程

1. 將主客場欄位改成 0, 1
2. 將 GAME\_ID 轉成整數
3. 對手做有序編碼(Ordinal coding)
4. 將 WL 做編碼

建立迴歸模型

挑出連續型特徵篩選得出 True 的變數與特徵工程所做變數

交叉驗證

與此同時分拆數據集, 設定 20%的訓練集合為驗證集合(Validation set), 並對訓練集合創建線性迴歸, 得到其特徵係數與截距項

評估修正模型

透過殘差圖、判定係數 R square、R square adj、MSE(損失函數)、MAE 來評估模型, 並去修正此模型(變數的多寡)

評估修正模型

最後得出最佳化迴歸模型, 並預測季後賽前 16 場的得分!