

統計學習初論 期末報告

MLB SALARY

(一頁 A4 描述程式碼與投影片中結果的關聯)

程式	片段	說明	對應投影片頁數
merge.R	ALL	用於將 MLB 官網和 SPOTRAC 的數據合併起來	8
merge.R	6	更改 year 變數來決定要合併的年份	(無對應)
blending_test.R	25~107	將所有年份的資料合起來，成為我們的 dataset	8
blending_test.R	117~134	不同等級（由高至低分成 20 等）的薪水平均畫圖，可發現成指數曲線	11
blending_test.R	137~160	不同站位的薪水 boxplot	12
blending_test.R	89~90	資料 feature 一覽，colnames 中 h_、f_、p_ 開頭分別代表對應 Hitting、Fielding、Pitching 類型的資料	13
cor_heatmap.R	ALL (126~132)	球員數據與明年薪水的 Correlation Heatmap，投影片中展示的是 fielder 的圖	14
cor_heatmap.R	ALL (156~176)	Correlation Heatmap 經過排序後，找出對不同位置而言，哪些數據影響明年薪水最多	15
blending_test.R	301~347	Chi2 Feature Selection	16
blending_test.R	78~82	Feature Normalization	17
blending_test.R	207~236	Regression	18
blending_test.R	350~387	Logistic Regression	24
blending_test.R	391~522	RandomForest	25
blending_test.R	526~623	Libsvm	26
blending_test.R	626~720	Xgboost	27
blending_test.R	257~270、291~297	資料分割、10 fold	28
blending_test.R	747~762	各個 Classifier 出來的平均準確度，站位請調整第 14 行的 pos_assigned	29~31
blending_test.R	374~381、506~515、609~617、709~713	各個 Classifier 的 testing 結果加上權重投票 (Ensemble)	33~34
blending_test.R	770~786	投票結果 (Ensemble Result)	36