



## D29：特徵評估



PDF 下載

全螢幕

## Sample Code & 作業內容

參考範例程式碼Day\_029\_Feature\_Importance.ipynb

先用梯度提升機對鐵達尼生存預測做訓練，再用其特徵重要性回答下列問題：

作業1：將特徵重要性較低的一半特徵刪除後，再做生存率預估，正確率是否有變化？

作業2：將特徵重要性最高的兩個特徵做特徵組合，是否能再進一步提升預測力？

作業請提交Day\_029\_HW.ipynb

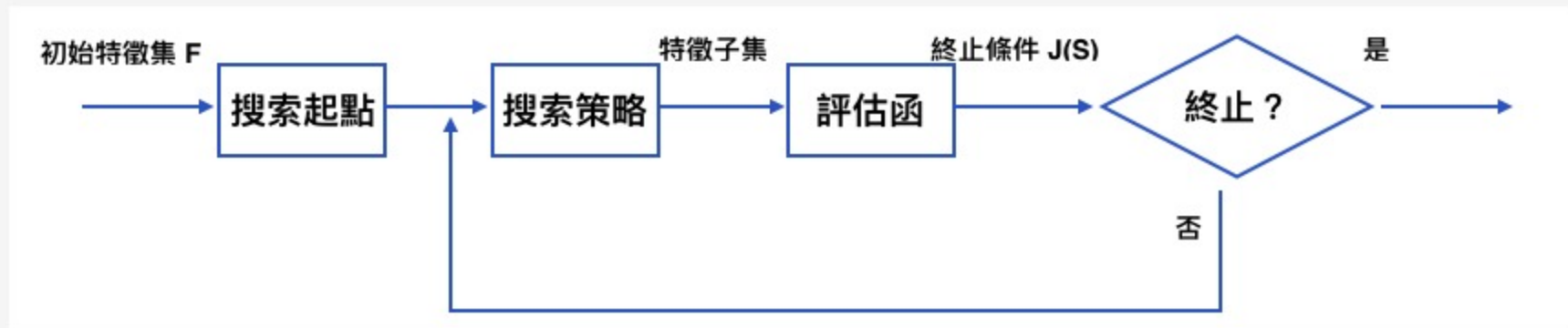
檢視範例

## 參考資料

### 機器學習 - 特徵選擇算法流程、分類、優化與發展綜述

掘金 [網頁連結](#)

有關特徵選擇的優化流程，在這邊有更完整的說明，不過這篇文章與其說是說明，不如說更像一份索引，我們可以在這篇文章中找到相當多的名稱與論文選錄，建議同學在專題 / 競賽當中遇到瓶頸時，不妨來逛逛這篇，尋找一下靈感。



### Permutation Importance

Kaggle Dan B. [網頁連結](#)

這裡是 Kaggle 上 Dan B. 提供的課程網頁，介紹我們課程中提到的排列重要性，雖然在樹狀模型上，其精準度略遜於特徵重要性，但是這個方法在非樹狀模型上也適用，泛用性不差。

Dan B. 在這系列課程中還有其他有趣的內容，例如 SHAP Value，能將樹狀模型預測的各個特徵影響性都可解釋化，在某些應用上，這個會比精準度還要有用。

Weight	Feature
0.0750 ± 0.1159	Goal Scored
0.0625 ± 0.0791	Corners
0.0437 ± 0.0500	Distance Covered (Kms)
0.0375 ± 0.0729	On-Target
0.0375 ± 0.0468	Free Kicks
0.0187 ± 0.0306	Blocked
0.0125 ± 0.0750	Pass Accuracy %
0.0125 ± 0.0500	Yellow Card
0.0063 ± 0.0468	Saves
0.0063 ± 0.0250	Offsides
0.0063 ± 0.1741	Off-Target
0.0000 ± 0.1046	Passes
0 ± 0.0000	Red
0 ± 0.0000	Yellow & Red
0 ± 0.0000	Goals in PSO
-0.0312 ± 0.0884	Fouls Committed
-0.0375 ± 0.0919	Attempts
-0.0500 ± 0.0500	Ball Possession %

## 提交作業

請將你的作業上傳至 Github，並貼上該網網址，完成作業提交

<https://github.com/>

確定提交

[如何提交](#) ▼

## 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](#) ▼