

D47：超參數調整與優化

### Sample Code & 作業內容

請使用不同的資料集，並使用 hyper-parameter search 的方式，看能不能找出最佳的超參數組合。

請參考範例程式碼Day\_047\_hyper\_parameter\_tunning.ipynb，作業請提交Day\_047\_HW.ipynb。

### 參考資料

* [**劍橋實驗室教你如何調參數 -  英文**](https://cambridgecoding.wordpress.com/2016/04/03/scanning-hyperspace-how-to-tune-machine-learning-models/)
* [**教你使用 Python 調整隨機森林參數 - 英文**](https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74)

## D48：Kaggle 競賽平台介紹

### Sample Code & 作業內容

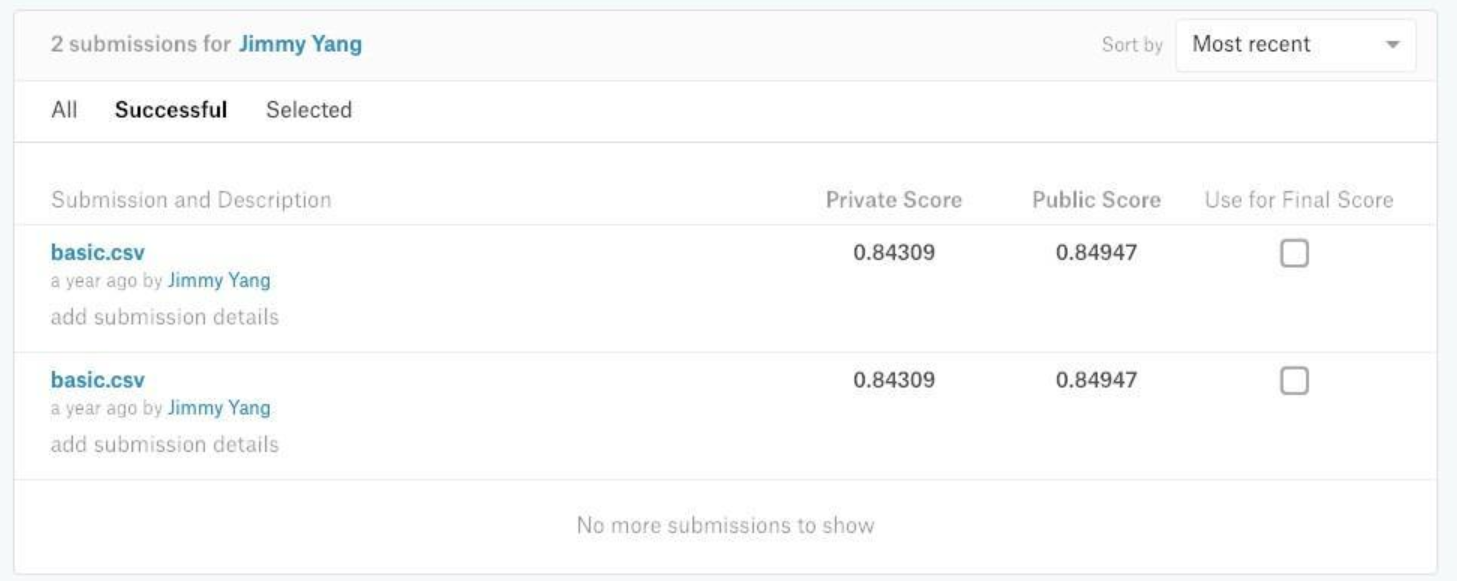
本日作業：

透過 Kaggle 的實戰練習，把資料下載下來，自己做訓練集/測試集的切分，送進模型訓練，評估結果，再把測試集的答案上傳到 Kaggle 得到最終成績。

請完成 **[scikit-learn-practice](https://www.kaggle.com/c/data-science-london-scikit-learn" \t "_blank)** 比賽(點擊連結可至競賽頁面)，讓大家熟悉 Scikit-learn 的比賽

* 總共有一千筆訓練資料、40個 features，二元分類問題，練習 features scaling、建模、調參數等步驟
* 每天最多上傳 10 次結果
* 請在 private / public leaderboard 上取得 0.7 以上的準確率
* 可多參考別人的 Kernel，學習別人的寫法與思路，完成自己的 Kaggle 競賽

作業提交請截圖kaggle競賽頁面提交畫面上傳至github，並回到官網提交github連結。(以下為Kaggle競賽頁面截圖範例)



## D49：集成方法 : 混合泛化(Blending)

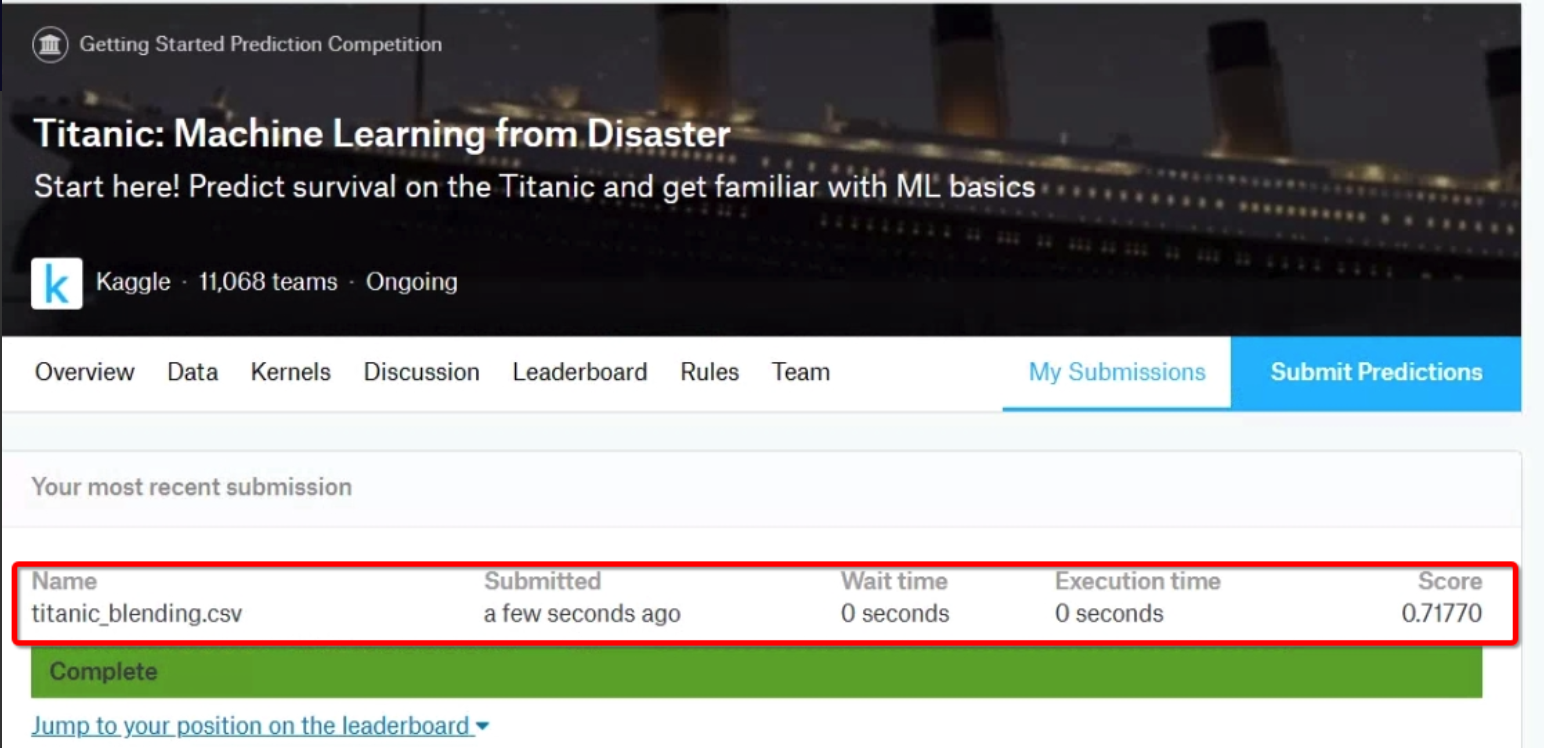
### Sample Code & 作業內容

雖然同樣是混合泛化，分類預測其實與回歸預測有相當多的差異性， 因為鐵達尼預測的結果是 '生存/死亡'，輸出不是 0 就是 1。因此要用權重混合時，需要以以機率的形式混合，因此我們在作業前幾格當中，先幫各位同學把預測值寫成了機率的形式

#### 今日作業

請參考範例程式碼Day\_049\_Blending，完成Day\_049\_Blending\_HW(作業檔)之後輸出成csv檔上傳到kaggle ，並將結果提交到 Kaggle 網站看看結果。(Kaggle鐵達尼競賽連結作業檔內提供)

作業提交請截圖kaggle競賽頁面提交畫面上傳至github，並回到官網提交github連結。(以下為Kaggle競賽頁面截圖範例)



#### 作業內容提醒

但是光是這樣，分類問題的混合泛化就能比單模預測還要好嗎?

已經快要期中考了，這裡請同學挑戰看看，還有沒有什麼方法可以改進混合泛化的結果?

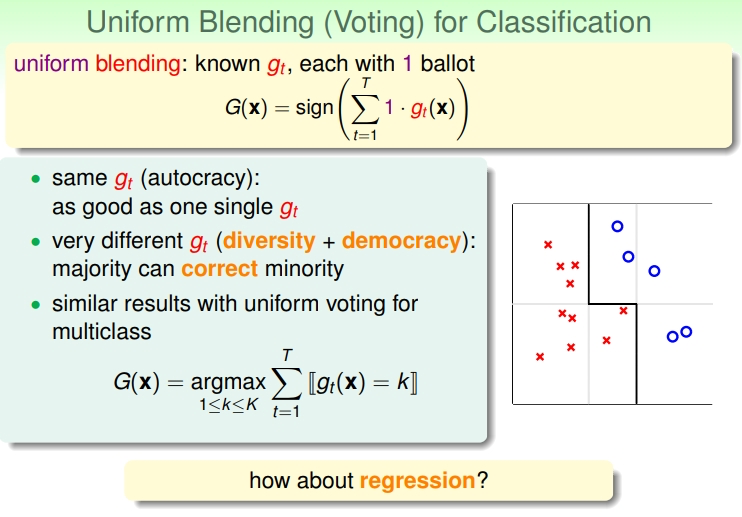
### 參考資料

#### 機器學習技法 Lecture 7: Blending and Bagging

林軒田老師公開課程  [**網頁連結**](https://www.csie.ntu.edu.tw/~htlin/mooc/doc/207_handout.pdf): <https://www.csie.ntu.edu.tw/~htlin/mooc/doc/207_handout.pdf>

[**影片連結**](https://www.youtube.com/watch?v=mjUKsp0MvMI&list=PLXVfgk9fNX2IQOYPmqjqWsNUFl2kpk1U2&index=27&t=0s): <https://www.youtube.com/watch?v=mjUKsp0MvMI&list=PLXVfgk9fNX2IQOYPmqjqWsNUFl2kpk1U2&index=27&t=0s>

當我們在網路上自己搜尋 Blending 時，往往搜尋到的都是林軒田老師的課程筆記，因此我們推薦同學如果對於 Blending 或  Bagging 的理論想要一探更完整內容的話，不妨來這邊尋找研讀的資料，相信絕對不會讓您失望 (如果太困難，也可以參考網路上眾多的閱讀筆記)

圖片來源：林軒田 機器學習技法 Lecture 7: Blending and Bagging

#### Superblend

Kaggle 競賽網站-Kernel 範例 [**網頁連結**](https://www.kaggle.com/tunguz/superblend)

這邊就是我們所謂競賽中的 Blending Kernel，只是決定一個權重，將兩個其他的 Kernel 合併成答案檔，就是這場競賽中的最高分 Kernel，我們並不是要鼓勵大家也去這樣去賺分數，而是在告訴大家 : Blending 的簡單，以及 Blending 的具有威力。



圖片來源：kaggle

## D49：集成方法 : 混合泛化(Blending)

### Sample Code & 作業內容

雖然同樣是混合泛化，分類預測其實與回歸預測有相當多的差異性， 因為鐵達尼預測的結果是 '生存/死亡'，輸出不是 0 就是 1。因此要用權重混合時，需要以以機率的形式混合，因此我們在作業前幾格當中，先幫各位同學把預測值寫成了機率的形式

#### 今日作業

請參考範例程式碼Day\_049\_Blending，完成Day\_049\_Blending\_HW(作業檔)之後輸出成csv檔上傳到kaggle ，並將結果提交到 Kaggle 網站看看結果。(Kaggle鐵達尼競賽連結作業檔內提供)

作業提交請截圖kaggle競賽頁面提交畫面上傳至github，並回到官網提交github連結。(以下為Kaggle競賽頁面截圖範例)

#### 作業內容提醒

但是光是這樣，分類問題的混合泛化就能比單模預測還要好嗎?

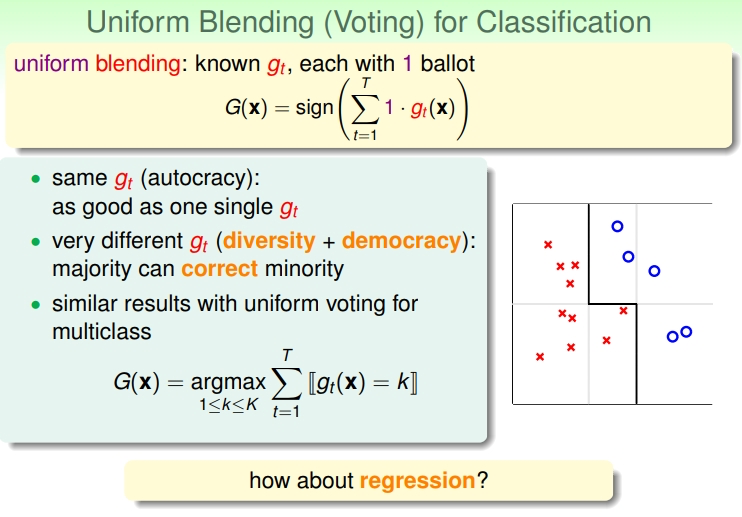
已經快要期中考了，這裡請同學挑戰看看，還有沒有什麼方法可以改進混合泛化的結果?

### 參考資料

#### 機器學習技法 Lecture 7: Blending and Bagging

林軒田老師公開課程  [**網頁連結**](https://www.csie.ntu.edu.tw/~htlin/mooc/doc/207_handout.pdf) [**影片連結**](https://www.youtube.com/watch?v=mjUKsp0MvMI&list=PLXVfgk9fNX2IQOYPmqjqWsNUFl2kpk1U2&index=27&t=0s)

當我們在網路上自己搜尋 Blending 時，往往搜尋到的都是林軒田老師的課程筆記，因此我們推薦同學如果對於 Blending 或  Bagging 的理論想要一探更完整內容的話，不妨來這邊尋找研讀的資料，相信絕對不會讓您失望 (如果太困難，也可以參考網路上眾多的閱讀筆記)

圖片來源：林軒田 機器學習技法 Lecture 7: Blending and Bagging

#### Superblend

Kaggle 競賽網站-Kernel 範例 [**網頁連結**](https://www.kaggle.com/tunguz/superblend)

這邊就是我們所謂競賽中的 Blending Kernel，只是決定一個權重，將兩個其他的 Kernel 合併成答案檔，就是這場競賽中的最高分 Kernel，我們並不是要鼓勵大家也去這樣去賺分數，而是在告訴大家 : Blending 的簡單，以及 Blending 的具有威力。

圖片來源：kaggle

Q&A

**依 Kaggle 傳回分數調整比重?**

請問老師：

依照範例的IN[14]

# 混合泛化預測檔 (依 Kaggle 傳回分數調整比重, 越準確者比重越高, 依資料性質有所不同)

這邊的比重不太懂是哪裡來的，謝謝。

A:

你今天訓練了A, B, C 三個模型，有三個答案可以上傳到 Kaggle，結果發現模型 A 的答案比較准一點，你就可以給模型 A 比較高的權重。

最終上傳的答案就會是 0.5\*A + 0.3\*B + 0.2\*C，當然權重的數值都可以根據你自己的判斷去調整囉

Q: **Bagging的使用情境**

您好，

在D45的**[完整的 Ensemble 概念 by 李宏毅教授](https://www.youtube.com/watch?v=tH9FH1DH5n0" \t "_blank)**  影片中(8:27~) ，李老師是說 bagging用在模型是complex , 容易 overfitting

但本日 D49說，Bagging是搭配弱分類器

請問這2種說法有何不同 ?

謝謝

A:

bagging 主要的想法是透過不同的模型組合使用，達到降低 variance 避免 overfitting 的情況，關於強弱分類器這件事倒不是那麼重要。

有興趣的話可以看一下這篇在知乎上的討論：[**https://www.zhihu.com/question/296676708**](https://www.zhihu.com/question/296676708)

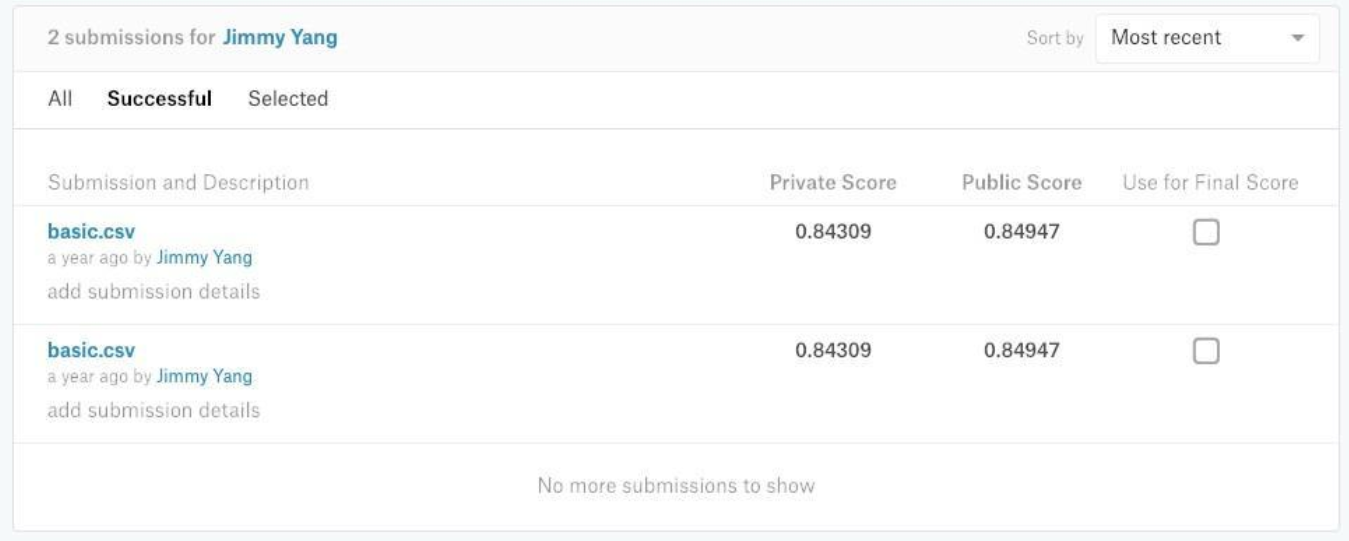
## D51-D53：Kaggle期中考

### Sample Code & 作業內容

請至以下Kaggle競賽平台期中考連結參加考試，請按join competition加入測驗

**並將任一次提交完成畫面截圖，上傳到github並回到官網完成提交。(提交畫面截圖範例)**

[**https://www.kaggle.com/t/2eaf9a8ed13847cd9c26f05b2abfe966**](https://www.kaggle.com/t/2eaf9a8ed13847cd9c26f05b2abfe966)



同學提交檔案前可參考Baseline Kernel作法

[**https://www.kaggle.com/ml100marathon/baseline-example**](https://www.kaggle.com/ml100marathon/baseline-example)