

## D49：集成方法：混合泛化(Blending)



PDF 下載

全螢幕

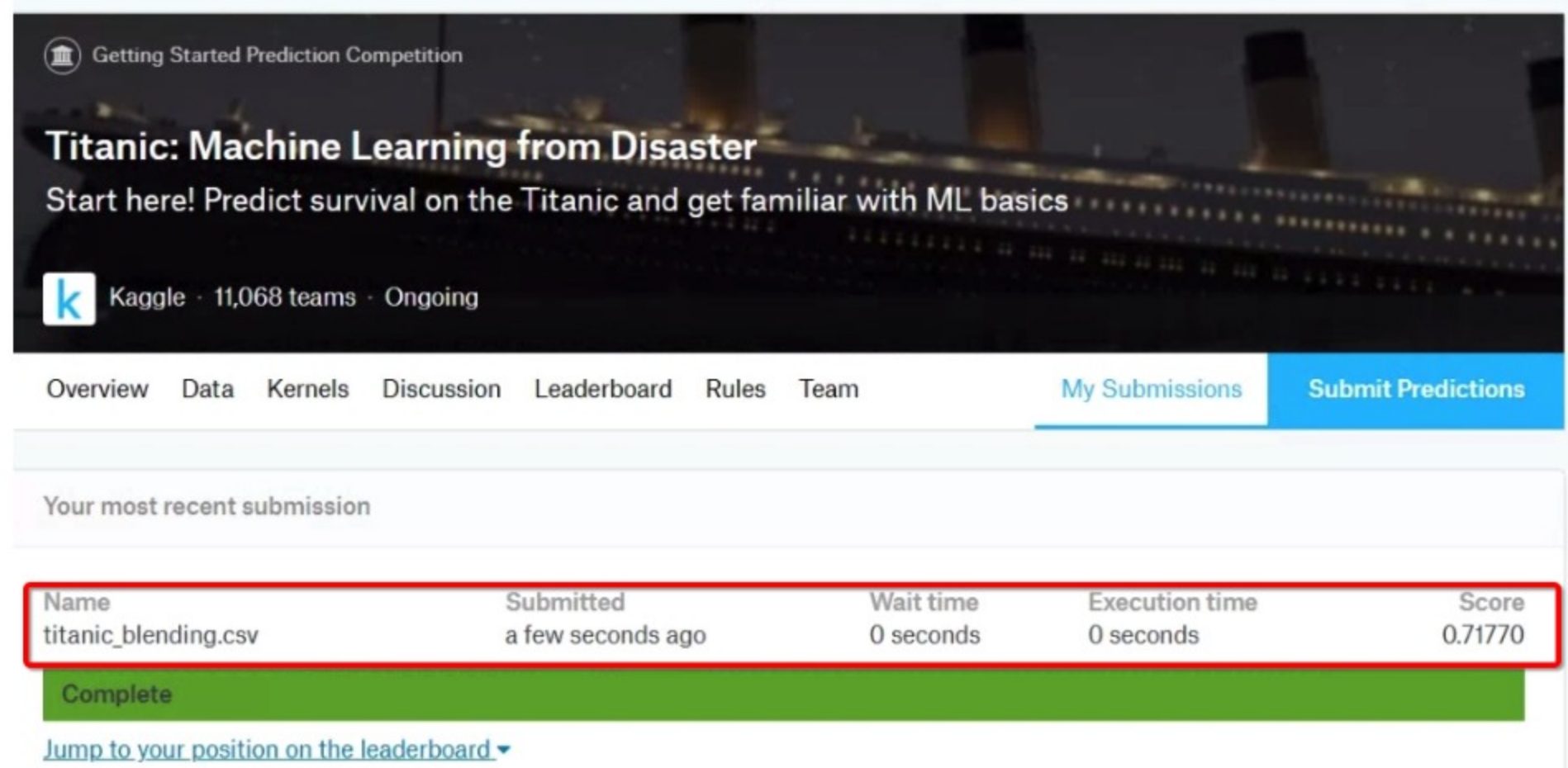
## Sample Code &amp; 作業內容

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？

檢視範例

## 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](#) [影片連結](#)

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)

### Uniform Blending (Voting) for Classification

uniform blending: known  $g_i$ , each with 1 ballot

$$G(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^T 1 \cdot g_i(\mathbf{x}) \right)$$

- same  $g_i$  (autocracy): as good as one single  $g_i$
- very different  $g_i$  (diversity + democracy): majority can correct minority
- similar results with uniform voting for multiclass

$$G(\mathbf{x}) = \underset{1 \leq k \leq K}{\text{argmax}} \sum_{i=1}^T \mathbb{I}[g_i(\mathbf{x}) = k]$$

how about regression?

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](#)

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

## 提交作業

請將你的作業上傳至 Github，並貼上該網址，完成作業提交

確定提交

[如何提交](#)

## 熱門問答

魏培峰 · 2019.06.08

## Bagging的使用情境

您好，在D45的完整的 Ensemble 概念 by 李宏毅教授 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

## 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](#)