

1. 請比較你實作的 generative model、logistic regression 的準確率，何者較佳？

	Public score	Private score
Logistic regression	0.85196	0.85333
Generative model	0.84619	0.84117

不論 public 或是 private，Logistic regression 都略勝一籌。

2. 請說明你實作的 best model，其訓練方式和準確率為何？

我使用 sklearn 裡的 gradient boost classifier 並且用 grid search 找出最好的 number of estimator 700，最後出來的 public score 為 0.87542、private score 為 0.87667。

3. 請實作輸入特徵標準化(feature normalization)並討論其對於你的模型準確率的影響

Unscaled / Scaled	Public score	Private score
Logistic regression	0.79852 / 0.85196	0.79474 / 0.85333
Generative model	0.76707 / 0.84619	0.76074 / 0.84117
Best	0.86977 / 0.86977	0.86537 / 0.86537

做標準化後所有的方法準確度都大量提升，可見資料標準化對於訓練 model 有所幫助。

4. 請實作 logistic regression 的正規化(regularization)，並討論其對於你的模型準確率的影響。

在沒有 regularization 的情況下，用原來的 logistic regression 的準確度為 0.85196 / 0.85333 (Public / Private)，而使用 regularization lambda = 0.01 後，準確度稍微提升到 0.85323 / 0.85501 (Public / Private)，影響並不是很大。

5. 請討論你認為哪個 attribute 對結果影響最大？

對於每個 attribute 假設一個值，如果值大於他就是 true，反之則為 False，再去跟 y 做比較，發現比較出來 captain gain 以及 captain loss 這兩個 attributes 的準確度最高，因此我認為 captain gain 以及 captain loss 影響最大。