

學號：B05902125 系級：資工三 姓名：葛洵為

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

Generative model: Private:0.84117 Public:0.84631

Logistic regression: Private:0.85284 Public:0.85245

很明顯的可以看到 Logistic 結果優於 Generative model

2. 請說明你實作的 **best model**，其訓練方式和準確率為何？

方法為 logistic regression，並沒有使用套件，把第 1,2,6 項的 feature 平方，然後做 normalize，使用 adagrad，iteration 為 10000 次

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

Without normalization: Private:0.79339 Public:0.79680

With normalization: Private:0.85284 Public:0.85245

對於有 normalization 可以猜測助於避免 overfit 的情形，以及猜測資料數字其實很大。

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。

Lamda = 100.0 Private:0.85173 Public:0.85356

Lamda = 10.0 Private:0.85517 Public:0.85700

Lamda = 1.0 Private:0.85702 Public:0.85675

Lamda = 0.1 Private:0.85726 Public:0.85687

可以看到 regularization 後再 private 上面的表現有顯著的提升，並且隨者 lamda 越低而越高

5. 請討論你認為哪個 **attribute** 對結果影響最大？

我認為 **age** 對結果影響最大，根據對 **age** 多次方以後不太影響 model 甚至更好。