

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

答：

Generative	0.84570
Logistic	0.85958

Logistic 比較好。

2.請說明你實作的 **best model**，其訓練方式和準確率為何？

答：

把一些連續的特徵加上多次項並 normalize 後，用 xgboost 的 binary logistic 訓練。Public score=0.87162。

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。

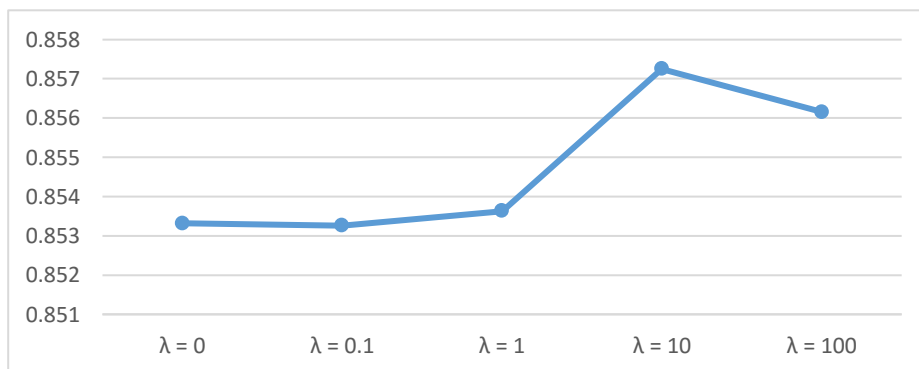
答：

Logistic	With normalization	0.85332
	Without normalization	0.23622
Generative	With normalization	0.76377
	Without normalization	0.84368

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。

答：

$\lambda = 0$	0.85332
$\lambda = 0.1$	0.85326
$\lambda = 1$	0.85363
$\lambda = 10$	0.85725
$\lambda = 100$	0.85615



正規化後準確率隨著  $\lambda$  的加大而提高，但是在  $\lambda=10$  後準確率又開始下降。

5.請討論你認為哪個 **attribute** 對結果影響最大？

我認為特徵的標準化對結果的影響最大，尤其是在 logistic 的做法中，若沒有讓 feature 間的標準一至，就容易 overflow