

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：

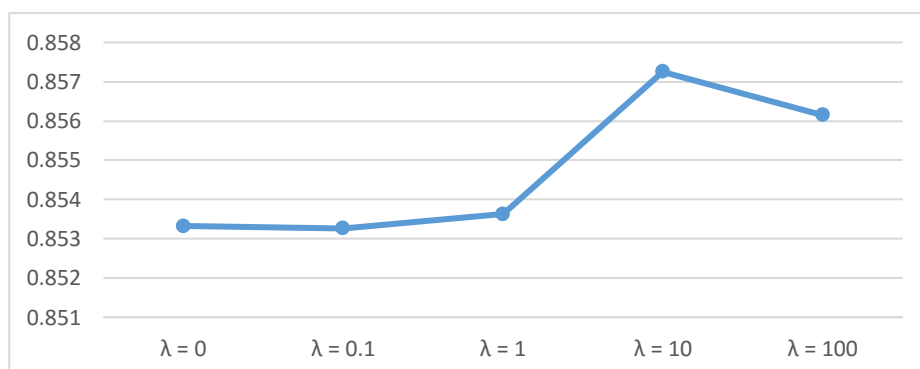
Logistic	With normalization	0.85332
	Without normalization	0.23622
Generative	With normalization	0.76377
	Without normalization	0.84368

在 **logistic** 中，**normalize** 後的結果比較理想。但在 **generative** 中反而直接使用原始資料會較準確，推測是因為 **normalize** 後會影響到 **generative function** 會使用到的 **mean** 和 **standard deviation**。

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？

根據實作，我認為 `capital_loss` 的影響最大，所以取了較高次。比較 1~5 次方的結果（根據 validation）如下：

```
1: 0.8570112400958172
2: 0.8569498188071986
3: 0.8571340826730545
4: 0.8572569252502917
5: 0.8571955039616731
```

所以我取了 `capital_loss` 的四次方。