

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

答：

	public	private	average
Generative	0.84484	0.84240	0.84362
Logistic	0.85773	0.85493	0.85633

Logistic 的效果比 Generative 好很多

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

(平均)準確率:0.85633

我是使用三階的 logistic model，並使用 regularization 避免 overfit (lamda=0.1)

如果只使用二階，準確率會略差，為 0.85602，加上 regularization 後也只有 0.85608

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：

如果不使用 normalization, (平均)準確率為:0.82372

加上 normalization 後, (平均)準確率為:0.85535 (其他參數定為相同)

所以有 normalize 的效果比較好，因此其他測試的 model 都是有加上 normalization

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。

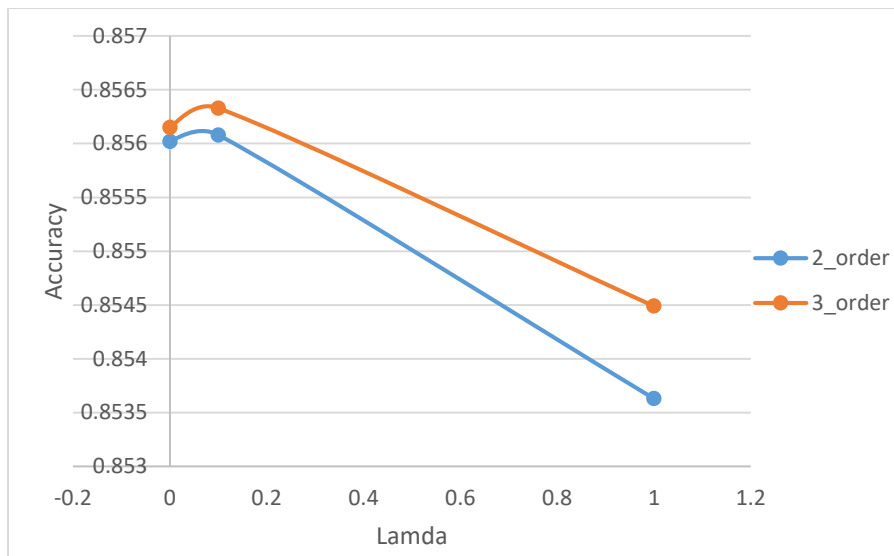
答：

(1) 使用二階的 logistic, iteration=3000

	Accuracy
Lamda=0	0.85406
Lamda=0.01	0.854
Lamda=0.1	0.85443
Lamda=0.2	0.85449
Lamda=1	0.85369
Lamda=10	0.84546

可以看出 lamda=0.2 時效果最好，但其實跟沒加 regularization 的時候(lamda=0) 差異不大，應該是因為我們的 model 還沒有到達 overfit，所以效果不顯著

(2) 比較二階級三階的 logistic, iteration=10000



因為 iteration 做到 10000 次很浪費時間，所以沒有測試太多的 lamda 值，依然可以看出當 lamda=0.1 時，效果比沒有加 regularization 好一點，但也只有好一點點而已，所以就沒有再更深入研究。

5.請討論你認為哪個 attribute 對結果影響最大？

Machine learning 就是機器會自己去學習各個 weight 最適當的值，當某個 attribute 對應到的 weight 越大時，代表這個 attribute 越重要。因此可以看出，capital\_gain 這項最重要。