學號：R06723025 系級： 財金碩一 姓名：林耘寬

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

	Public	Private
Generative	0.84557	0.84215
Logistic	0.86044	0.85702

由圖表可知 Logistic model 較佳。對於 Generative model 而言，我只有單純使用助教提供的 feature 下去算 mean, covariance matrix，但 Logistic regression 我做了相當多嘗試，也是我的 best model，詳情如第二題所說，所以 Logistic regression 呈現出來的結果也比較好。

2.請說明你實作的 best model，其訓練方式和準確率為何？
答：
Best model 即是用 logistic regression 做的。其中對 Column 0, 1, 3, 4 ,5 加上高次方項做為新的 feature。我將前百分之十的 training set 設成 validation set，然後依序從一次方加到十次方項，可以觀察到在六次方項後雖然對於 training set 的正確率較好，但對於 validation set 卻呈現正確率下降的趨勢，故推測有 overfitting 的現象，而且再三次方的時候，對於 validation set 的正確率最高，所以採用三次方以及六次方訓練出來的結果作為最後的選擇，但結果發現對於 private set 而言，二次方的結果反而有最好的結果，推測可能有 overfitting 的現象。

	Public	Private
一次方	0.85012	0.84166
二次方	0.86019	0.85726
三次方	0.86031	0.85603
六次方	0.86044	0.85702

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 regression 在未經過 normalize 的情況下，使用 learning rate = 0.1，會在訓練過程產生 overflow 的現象。所以我們直接嘗試特徵標準化，同樣在 learning rate = 0.1 的情況下，使用助教是先抽取的全部 feature，就得到了第一題以及第二題的結果。

Generative model 的結果如下表所示，可以觀察到差異並不顯著。

Generative model	Public	Private
With normalization	0.84557	0.84215
Without normalization	0.84520	0.84240

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	Public	Private
0	0.86019	0.85726
0.1	0.86031	0.85726
0.01	0.86019	0.85726

可以發現影響不大，上述的模型軍只有採用到二次項，故可能較無 overfitting 的問題。

5.請討論你認為哪個 attribute 對結果影響最大？
針對這個問題，我們只針對第 0, 1, 3, 4, 5 個 feature 來處理，因為其他的數據多數為 0, 1，其中零為大多數。處理的方法為將其他 feature 依序加上上述某個 feature 觀察他們的正確率，發現加入第三個 feature 會使得正確率上升大約 10%，其餘較無影響。所以我認為第三個 feature- capital gain 對結果影響較大。

加入的參數	0	1	3	4	5
正確率	0.72182	0.71793	0.82767	0.72161	0.72120