學號：R06943153 系級： 電子碩一 姓名：蘇旻彥

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

	Kaggle public score	Kaggle private score
generative	0.84606	0.84252
logistic	0.85393	0.85124

由上表可知 logistic regression 的結果比較好一些。

2.請說明你實作的 best model，其訓練方式和準確率為何？
答：

這次作業我所選的 best model 是使用 logistic 的 model，使用了 Adagram，並對 input 加上 square term 之後的結果，在 Kaggle 上面的 public score 得到了 0.85921 的分數。

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：

使用第二題的 best model 作比較，把 normalization 拔掉之後的結果如下

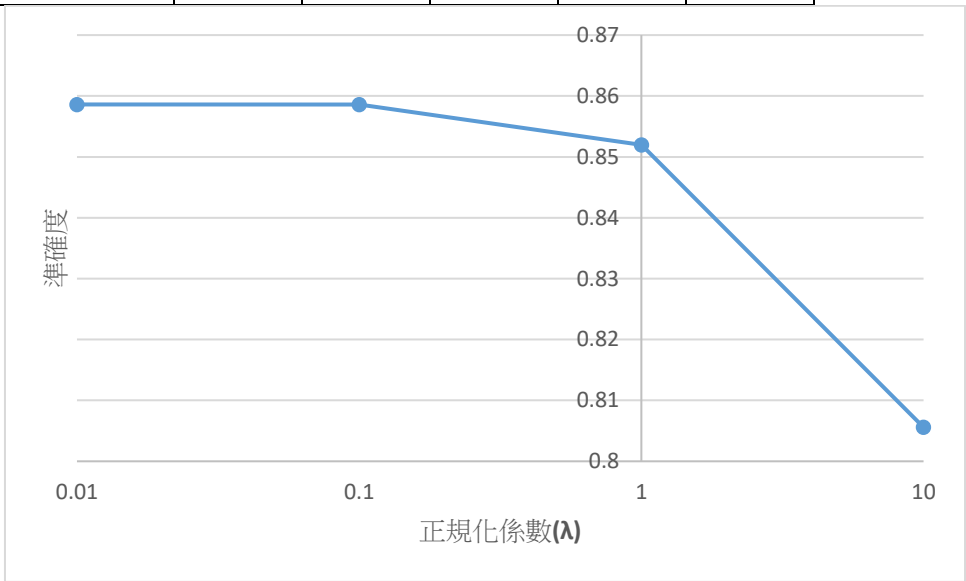
	Kaggle public score	Kaggle private score
With normalization	0.85921	0.85566
Without normalization	0.78918	0.78663

明顯的可以看出沒有作normalization的結果會變得非常差。我覺得是因為原本特徵內的數字可能過大，經過sigmoid函數後差距變小(轉會後都很靠近1)，而看不出其中的差別，使結果變差。

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：

為了方便這題使用的是 valid score 而不是 Kaggle 上面的分數。由下表可以看到當家入正規化之後，隨著正規化係數(λ)的增大，準確率會微微的下降，但差異並不會很大；當正規化係數大於 1 之後準確度會急遽的下降。

λ	0	0.01	0.1	1	10
valid score	0.8576	0.8586	0.8586	0.852	0.8056



5.請討論你認為哪個 attribute 對結果影響最大？

答：

在加入平方項之前，取出所有特徵共 106 個 weights 來看，超過 0.2 的只有 15 個，但光取這 15 個再做一次，結果並沒有想像中的突破性成長。接著嘗試加入平方項，因為有點懶惰所以連 onehot 的 feature 也一並作了平方項目，所以共有 212 個 features。這 212 個當中最後結果 weight 有超過 0.2 的，去掉有些 onehot 的平方之後取出前 11 名，按照分數排列如下：

id	name	weight
0	age	3.181171
3	capital_gain	2.315249
5	hours_per_week	0.908491
109	capital_gain**2	0.678827
33	Married-civ-spouse	0.371717
24	Bachelors	0.264363
110	capital_loss**2	0.238833
41	Exec-managerial	0.234441
54	Not-in-family	0.213704
2	sex	0.208798