

學號：B03705049 系級：資管四 姓名：廖寬璿

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

	public	private
generative	84.385	0.84117
logistic	84.595	0.84178

這兩個準確率差不多logistic稍微好一點點

2.請說明你實作的best model，其訓練方式和準確率為何？

答：

	public	private
xgboost	87.579	0.87126

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。

	public	private
baseline	0.76523	0.76231
normalized	0.65945	0.66416

對這個模型(&沒選feature)normalize後的結果變差
但在其他組合normalize有變好

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。

	public	private
baseline	0.76523	0.76231
regularized	0.76523	0.76231

feature regularization沒影響

5.請討論你認為哪個attribute對結果影響最大？

用sklearn 的random forest: feature importance 最高的是capital gain, 再來是age