

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

答：

generative model的準確率較佳

	Public Score	Private Score
generative model	0.84533	0.84252
logistic regression	0.83009	0.81869

2.請說明你實作的best model，其訓練方式和準確率為何？

答：

利用 sklearn 裡的package：GradientBoostingClassifier，

tree_num = 1000，learning rate = 1

其訓練方式是搭配decision tree及Gradient decent的方式來training

	Public Score	Private Score
GradientBoostingClassifier	0.86977	0.86500

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：

從Score來看，無論是哪個model，做過normalization後的Score都是比較好的；

從Model來看，在generative model的部份影響不大，我猜應該是因為我的generative model是用gaussian distribution來計算，原本就是以整體來看待，而logistic的部份提昇蠻顯著的，可能是因為做了normalize後，bias也跟著被normalize了

	Public Score	Private Score
generative	0.84520	0.84240
generative_norm	0.84533	0.84252
logistic regression	0.79471	0.79093
logistic regression_norm	0.83009	0.81869

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：

做了regularization之後，對於testing set的結果有微幅上升，可以猜想是沒做regularization時某些data使得model overfitting，而做完 regularization 之後使得曲線平滑，讓overfitting的現象平緩了。

5.請討論你認為哪個attribute對結果影響最大？

我認為是capital_gain最有關聯，因為年收入較高者才有額外的資金去做投資，像是股票、土地等等投資