

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

答：

	Private Score	Public Score
Generative Model	0.84227	0.84520
Logistic Regression	0.84964	0.85552

Logistic Regression 的準確率較佳。

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

Best Model 為使用 gradient boosting 的方式來實作，訓練方式為使用 Scikit-learn 0.19.0 套件中的 gradient boosting 來實作，並且設定 depth=5。

	Private Score	Public Score
Gradient Boosting	0.87372	0.87874

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：

	標準化前	標準化後
Generative Model	0.84227	0.84227
Logistic Regression	Overflow	0.84964
Gradient Boosting	0.87372	0.87372

Generative Model - 有沒有標準差都沒差別。

Logistic Model - 沒作標準化會 overflow 無法實作。

Gradient Boosting - 有沒有標準差都沒差別。

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：

	$\lambda = 100$	$\lambda = 10$	$\lambda = 1$	$\lambda = 0.1$	$\lambda = 0.01$
Accuracy	0.84166	0.84166	0.84160	0.85057	0.80650

在 $\lambda = 0.1$ 時會較好

5.請討論你認為哪個 attribute 對結果影響最大？

根據 sigmoid 與 likeness function，w 絕對值越大的 attribute 影響越大。

capital_gain 越高，年收越會高於 50K。

工作經驗是 Never-worked 的，年收越會低於 50K。

Without-pay 越高，年收越會低於 50K。

學歷是 Preschool 的年收越會低於 50K。

住在 Holand-Netherlands 的年收越會低於 50K。