

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

答：Generative: 0.84240

Logistic: 0.84350

Logistic 略優

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：Sklearn 中的 AdaBoostClassifier (n_estimators=10000)

Score: 0.87114

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：

	With normalization	Without normalization
Logistic	0.84350	0.78012
Generative	0.84215	0.84240

Logistic 在有標準化的情況下正確率提升相當高，Generative 則幾乎無影響

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：

	$\lambda=0$	$\lambda=0.001$	$\lambda=0.0001$
Logistic	0.84350	0.84092	0.84203

Regularization 對於正確率沒有提升

5.請討論你認為哪個 attribute 對結果影響最大？

Without	Logistic (all attribute = 0.84595)
Age	0.84706
Sex	0.84601
Capital gain	0.83268
Capital loss	0.84294
Hours per week	0.84564
Workclass	0.84425
Education	0.83784
Marital status	0.84583
Relationship	0.84503
Marital status & Relationship	0.82476
Occupation	0.84300
Country	0.84570

以上為分別拿掉某項 attribute 的正確率，可以看出 capital gain 和 education 影響最大，其中 marital status 與 relationship 都是人際相關，若是一起拿掉的話正確率則會掉最多