

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

答：

我的logistic regression使用了adagrad及對data做了normalization，train了1000次後，得到的public準確率為0.85417，private準確率為0.85087。

而generative model對feature做normalization，運算後，得到的public準確率為0.84557，private準確率為0.84227

可看出logistic regression的結果較佳

2.請說明你實作的best model，其訓練方式和準確率為何？

答：

我的best model是以xgboost套件train出來的，試過每顆tree的depth以3和6，train 20 round，並以depth為6的準確率較高。public準確率為：0.87371 private準確率為：0.86881，較logistic regression高出接近1.5%左右的準確率。

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：

基本上，logistic regression不做normalize，算出來的z值(丟進sigmoid function的值)會過大，以至於算出來的結果相當不準確(sigmoid出來的值全為0)，在經過對全部feature normalize後，logistic regression的predict才會準確。

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：

下表為做完regularization後的public與private成績發現準確度反而略微降低了，但並沒有明顯的影響

λ	0	0.1	0.01
public	0.85417	0.85403	0.85412
private	0.85087	0.85066	0.85083

5.請討論你認為哪個attribute對結果影響最大？

下表是logistic regression中，weight前五大的feature

No.	1	2	3	4	5
feature	capital_gain	Holand-Netherl ands	Never-worked	Preschool	Outlying-US(G uam-USVI-etc)
weight	2.37993488	-0.96437995	0.75782911	-0.73613856	-0.72478241

可看出capital_gain的weight明顯大於其他feature，可推測，capital_gain的attribute最大