

學號：R06922093 系級：資工碩一 姓名：陳禹齊

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

答：

	public	private
Generative model	0.76523	0.76231
Logistic regression	0.85331	0.85136

A: 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：

	public	private
Best model	0.85847	0.85345

說明: 使用 logistic regression,然後加上 adagrad, feature 取一次項加二次項, iteration 設 10000 次,learning rate 設 10。

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。

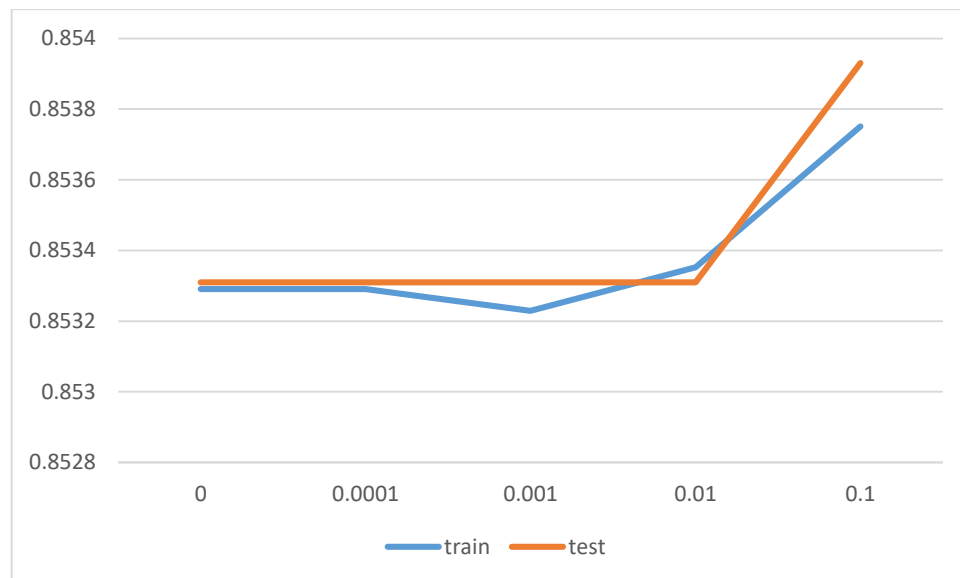
答：

	public	private
normalization	0.85331	0.85136
Without normalization	0.48624	0.48937

討論: 如果沒做標準化,前面幾項影響太大,所以只要前面幾項的值太大就會讓整體的表現很差,因為後面的項沒辦法影響太多。

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：



討論：在 0.1 的時候會有最好的表現,其他 lambda 的值都沒有比較好的表現。

5.請討論你認為哪個 attribute 對結果影響最大？

	train	差值(去掉項的值-正常值)
正常	0.83007279	0
去掉 age	0.82942784	-0.00064495
去掉 fnlwgt	0.83074844	0.00067565
去掉 sex	0.83464881	0.00457602
去掉 capital_gain	0.81247505	-0.01759774
去掉 capital_loss	0.82589601	-0.00417678
去掉 hours_per_weeks	0.82866005	-0.00141274

討論：根據差值來看,去掉 capital\_gain 後,準確率下降最多,所以我覺得 capital\_gain 對結果的影響最大。