學號：B06901087 系級：電機二 姓名：翁瑋襄

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

	public acc	private acc
generative model	0.84092	0.84680
logistic model	0.85345	0.85171

logistic model 訓練出來的準確度較佳。

2. 請說明你實作的best model，其訓練方式和準確率為何？

我的best model訓練方式是使用sklearn.emsemble.GradientBoostingClassifier，網路上查到他的實作方法是使用desicion tree，其預測結果在Kaggle的表現: public acc:0.87360, private acc:0.87567，結果相較於DNN來說更準確，我認為可能原因是本次作業的題目資料以one-hot encoding表示，較適合以decision tree做為model。

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

	with feature normalization	without feature normalization
public acc	0.85333	0.80149
private acc	0.85184	0.80343

有沒有做feature normalization差異真的滿大的。我認為可能的原因是，沒有做normalization前可能會有某一項變化特別的劇烈，進而影響到其他變化比較平緩的項次gradient的更新。

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。

lambda	1000	100	10	1	0.1	0.01	0.001	0
public/ private acc	0.84289 / 0.84815	0.85149 / 0.85270	0.85333 / 0.85184	0.85345 / 0.85159	0.85345 / 0.85159	0.85223 / 0.85147	0.85024 / 0.85136	0.84729 / 0.84817

從上表可以看出，加上regularization後，使得loss function變的平滑一些，對於train模型的準確率的確有提升，不過lambda也不可設太大，否則準確率又會望下掉。

5. 請討論你認為哪個attribute 對結果影響最大？

經過normalization後，我去找w裡面所有元素取絕對值後最大的那一項，發現是capital gain這一項，我認為還滿合理的，因為資本獲益量跟金錢的關係最直接，應該也會直接影響收入的多寡。