學號：B05502145 系級：電機三 姓名：林禹丞

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

	private	public	average
Generative	0.85222	0.8554	0.85381
Discriminative	0.85775	0.86142	0.859585

Discriminative 較佳。

2. 請說明你實作的 best model，其訓練方式和準確率為何？

使用 discriminative model。對 age、fnlwgt、caploss、capgain 加入高次項。Regularize 0.01。Iteration 4000 次。

	private	public	average
Discriminative	0.85775	0.86142	0.859585

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

	private	public	average
nonorm	0.78012	0.77948	0.7798
dis	0.85775	0.86142	0.859585

沒有作 normalization 的情況下，model 會 train 不起來。所以在 test 上的表現也很差。

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。

	private	public	average
reg1	0.85751	0.86179	0.85965
reg0.1	0.85787	0.86179	0.85983
reg0.01	0.85775	0.86142	0.859585
reg0.001	0.85837	0.86093	0.85965

基本上沒有什麼差別，可以防止 overfitting 的狀況。

5. 請討論你認為哪個 attribute 對結果影響最大？

	private	public	average
dis	0.85775	0.86142	0.859585
delsex	0.85259	0.85307	0.85283
delnone	0.85345	0.85171	0.85258
delhours	0.8494	0.85307	0.851235
delfnt	0.85173	0.85098	0.851355
delcaploss	0.8478	0.85122	0.84951
delcapgain	0.83417	0.84066	0.837415

delage	0.8513	0.85184	0.85157

分別刪掉不同的項目去比較準確率，可以發現刪掉 captain gain 之後，model train 不好。所以 captain gain 對這個問題的影響較大。