學號：B04902011　系級：資工三　姓名：張立暐

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

   註 1：用來訓練的 data 為助教提供的 106 項 train feature 各取一次項，並有將 train & test data 一起做標準化。
   註 2：這裡實作的 logistic regression 中，(batch_size, epoch) = (32, 1000)，且採用 Adam 與其參數為 (lr=0.001, β_1=0.9, β_2=0.999, ε=1e-8)。

Model	Training Accuracy	Total Testing Accuracy
generative model	0.84223	0.84393
logistic regression	0.85277	0.85314

   由表可看出 logistic regression 比起 generative model 在 testing 結果上有較好的表現。

2. **請說明你實作的 best model，其訓練方式和準確率為何？**

   這邊使用套件 XGBoost 來實作 binary classification (objective='binary:logistic')，訓練資料為助教提供的 train feature 共 106 項且有標準化，XGBoost.train 的各項參數為 (lr=0.03, lambda=3, seed=1126, num_boost_round=1000) 。
   此 model 所獲得的預測準確率為 0.87605。

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**

   註：這裡使用的 data, model 及其參數皆與第 1. 小題相同（標準化除外）。

Model	有 Normalization		無 Normalization	
	Training	Testing	Training	Testing
generative model	0.84223	0.84393	0.84257	0.84381
logistic regression	0.85277	0.85314	0.77943	0.77765

   由表可以看出有無 normalization 對於 generative model 的預測結果來說影響不大；而對於 logistic regression 則有非常顯著的影響，這可能是因為在無標準化的情況下，有些 feature 值的大小相差很多，因此被特定 feature 影響較大導致最後預測結果變差。

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。**

   註：這裡使用的 data, model 及其參數皆與第 1. 小題相同（正規化除外）。

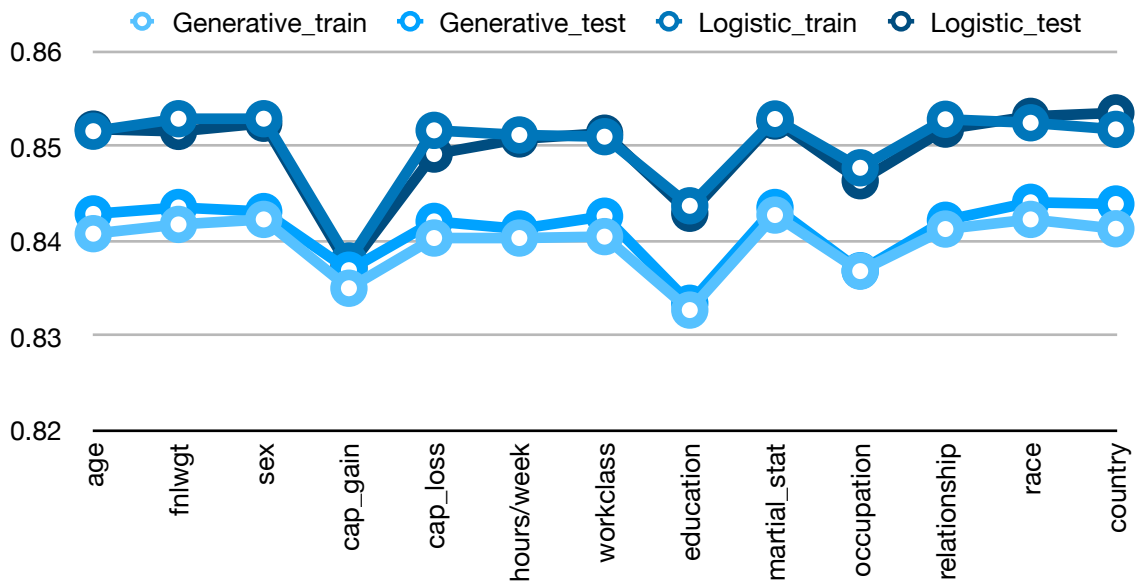
Lambda	Training Accuracy	Total Testing Accuracy
1	0.84423	0.84651
0.1	0.85243	0.85197
0.01	0.85274	0.85277
0.001	0.85277	0.85320
0.0001	0.85277	0.85320

由表可知，當我們將 lambda 調高時（增加正規化的影響），我們所得到的預測結果有下降的趨勢，所以在這個實作的情形下正規化並沒有帶來更好的效果。

5. **請討論你認為哪個 attribute 對結果影響最大？**

註：這裡使用的 data, model 及其參數皆與第 1. 小題相同 (Removed Attribute 除外)。

Removed Attribute	Generative Model		Logistic Regression	
	Training	Testing	Training	Testing
age	0.84079	0.84288	0.85163	0.85185
fnlwgt	0.84177	0.84356	0.85295	0.85161
sex	0.84230	0.84313	0.85295	0.85247
capital_gain	0.83502	0.83693	0.83790	0.83723
capital_loss	0.84033	0.84209	0.85172	0.84921
hours_per_week	0.84033	0.84129	0.85123	0.85081
workclass	0.84048	0.84264	0.85105	0.85142
education	0.83271	0.83343	0.84374	0.84295
martial_status	0.84276	0.84356	0.85292	0.85271
occupation	0.83683	0.83687	0.84776	0.84639
relationship	0.84128	0.84221	0.85289	0.85185
race	0.84230	0.84411	0.85252	0.85320
native_country	0.84131	0.84393	0.85182	0.85357



由上表發現，當我們把 capital_gain 從 attribute 中拿掉時，預測結果變差最多，判斷其可能為影響結果最大的 attribute，而拿掉 education 會有第二大的影響。