學號：B05705006 系級： 資管四 姓名：李和維

1. (0.5%) 請比較你實作的generative model、logistic regression 的準確率，何者較佳?

資料前處理皆為把 native_country 屬性移除。

	Generative Model	Logistic Regression
Public accuracy	0.84398	0.85601
Private accuracy	0.84412	0.85001

無論在 Public 或 Private leaderboard 上準確率都是 Logistic regression 比較好。

2. (0.5%) 請實作特徵標準化(feature normalization)並討論其對於你的模型準確率的影響

	Generative without normalization	Generative with normalization	Logistic without normalization	Logistic with normalization
Public accuracy	0.84398	0.82383	0.80098	0.85601
Private accuracy	0.84412	0.81660	0.79769	0.85001

標準化方法採用 Z-score 標準化。
可以觀察到對於 Logistic regreesion 來說，將數據標準化能讓準確率大幅提升，也較好調校 gradient descent 時的 learning rate。然而，對 Generative model 來說把數據標準化後準確率反而降低。
另外，best.py 中的 Gradient Boosting Classifier 準確率則是不受數據標準化影響。

3. (1%) 請說明你實作的best model，其訓練方式和準確率為何?
實作中得到準確率最高的 best model 是使用了 skikit-leran 的 Gradient Boosting Classifier。
訓練方式：

```
clf = GradientBoostingClassifier(n_estimators=250,
learning_rate=0.1, random_state=42, min_samples_split=200,
min_samples_leaf=50, max_depth=8, max_features='sqrt',
subsample=0.8).fit(X_train, Y_train.ravel())
```

調整 hyper parameters 的方式為，先固定一個較小的 n_estimators 與較大的 learning_rate，然後依據影響力高低調整其餘 hyper parameters（依序為 max_depth -> min_samples_split -> min_samples_leaf）。獲得較好參數後再開始調高 n_estimators 及調低 learning rate 增加模型 robustness。
準確率：

Public Accuracy	Private Accuracy
0.87592	0.86684

4. (3%) Refer to math problem

1、

$$L(\theta) = \prod_{n=1}^{N} \prod_{k=1}^{K} (P(x_n|C_k)\pi_k)^{t_{nk}}$$

$$\Rightarrow l(\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk}[\log P(x_n|C_k) + \log \pi_k]$$

We want to maximize $l(\theta)$ subject to $\sum_{k=1}^{K} \pi_k = 1$, hence introduce Lagrange Multipliers.

$$L(\pi,\lambda) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk}[\log P(x_n|C_k) + \log \pi_k] + \lambda(\sum_{k=1}^{K} C_k - 1)$$

$$\frac{\partial L}{\partial \pi_k} = \frac{1}{\pi_k}\sum_{n=1}^{N} t_{nk} + \lambda = 0 \Rightarrow \pi_k = -\frac{N_k}{\lambda}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{k=1}^{K} \pi_k - 1 \Rightarrow \sum_{k=1}^{K} \pi_k = 1$$

plug in all values into constraint:

$$\sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} -\frac{N_k}{\lambda} = -\frac{N}{\lambda} = 1 \Rightarrow \lambda = -N$$

$$\Rightarrow \pi_k = \frac{N_k}{N}$$

No.

Date

2、

$$\frac{\partial \log(\det\Sigma)}{\partial \sigma_{ij}} = \frac{1}{\det(\Sigma)} \frac{\partial \det\Sigma}{\partial \sigma_{ij}} = \frac{1}{\det\Sigma}(-1)^{i+j} M_{ij}$$

$$||$$

$$e_j \Sigma^{-1} e_i^T = e_j \frac{\tilde{\Sigma}}{\det\Sigma} e_i^T = \frac{1}{\det\Sigma}(-1)^{i+j} M_{ij}$$

3、 $$\frac{\partial L(\mu, \Sigma | x^n)}{\partial \mu_k} = \sum_{n=1}^{N} t_{nk} \Sigma^{-1} (\mu_k - x^n) = 0$$

$$\Rightarrow \underline{\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} t_{nk} x_n}$$

$$L(\mu, \Sigma | x^n) = C - \frac{N}{2} \log|\Sigma| - \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} \left[t_{nk} (x^n - \mu_k)^T \Sigma^{-1} (x^n - \mu_k) \right]$$

$$= C + \frac{N}{2} \log|\Sigma|^{-1} - \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} tr\left[t_{nk} (x^n - \mu_k)(x^n - \mu_k)^T \Sigma^{-1} \right],$$

$$\frac{\partial L}{\partial \Sigma^{-1}} = \frac{N}{2} \Sigma - \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} (x_n - \mu_k)(x_n - \mu_k)^T = 0$$

$$\xRightarrow{\Sigma = \Sigma^T} \Sigma = \frac{1}{N} \frac{N_k}{N_k} \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} (x_n - \mu_k)(x_n - \mu_k)^T$$

$$\Rightarrow \Sigma = \sum_{k=1}^{K} \frac{N_k}{N} S_k$$