學號：R06922117 系級：資工碩一 姓名：李岳庭

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

答：

	public	private	total
generative	0.85503	0.85358	0.85430530763
logistic	0.85540	0.85075	0.85307816831

feature 相同的情況下，generative model 表現較佳。

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

使用 scikit-learn 內的 AdaBoostClassifier()，將所有 training data 及 feature 丟進去，出來的成果是 public = 0.86302 、 private = 0.86021，我有試過將 feature 做正規化，但結果反而更差，我推論是要讓 feature 的差異越大，分類的效果越佳。

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：

normalization	public	private	total
no	0.80110	0.79633	0.79871856085
yes	0.85503	0.85358	0.85430530763

在做 logistic regression 的時候，對 feature 做 normalization 會有較高的準確度，而 normalization 是用以下公式：

$$x' = \frac{x - \bar{x}}{\sigma}, \quad \bar{x}: mean \ ; \ \sigma: standard\ deviation$$

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：

lambda	accuracy
0	0.84540
0.1	0.84540
1	0.84422
10	0.83030

Epoch 次數相同的情況下，lambda 越大 w 更新的幅度越低，所以準確度越低。

5.請討論你認為哪個 attribute 對結果影響最大？

答：

我沒有個別將每個 feature 去測試，我將幾個 continuous 的 features 的二次項加入 logistic model，對準確度有明顯的提升，而有做 normalization 也比沒做的準確度來得高。但在 scikit-learn 的套件下，使用 Decision Tree 或 AdaBoostClassifier 之類的分類器，nomalization 後反而使準確度下降。