學號：R06921082 系級： 電機碩一 姓名：陳與賢

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

上傳kaggle後分數分別如下
generative: **0.74164(public) / 0.73848(private)**
logistic: **0.85810(public) / 0.85444(private)**
logistic的準確率較佳

2.請說明你實作的best model，其訓練方式和準確率為何？
答：

其實我的best model就是logistic model，使用logistic加上gradient descent來實作，使用全部的feature額外加上年齡、capital_gain的平方項以及年齡的三次方項，因我覺得這些有高度相關，learning rate設定0.0001、訓練100000次，最後準確率為**0.85810(public) / 0.85444(private)**

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：

```
39 # data normalization
40 row_name = [0, 1, 3, 4, 5]
41 for i in row_name:
42     tmp = numpy.array(y_item[i])
43     print(numpy.amax(tmp))
44     tmp = tmp / numpy.amax(tmp)
45     y_item[i] = tmp.tolist()
```

Rescaling(除最大數值)來讓feature的值都介在0與1之間，會發現準確率提升許多。

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：

對於我的model的準確率沒有明顯提升，我覺得可能是因為原本的model就沒有overfitting的問題，所以 regularization並沒有太大功效，或是說我的learning rate設太小（但我怕設太大會underfitting）

5.請討論你認為哪個attribute對結果影響最大？
答：

我覺得年齡影響蠻大的，我原本只有使用所有feature的一次項，後來加上年齡的平方及三次方項之後準確率提升不少，我想一般而言，收入跟年齡成正比是一個比較普遍的現象。