學號：B03902125 系級： 資工四 姓名：林映廷

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

答：

logistic regression:(0.84903*8140+0.85307*8141)/16281=0.85105

generative model:(0.79719*8140+0.80147*8141)/16281=0.79933

在我的實作中，logistic regression 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：

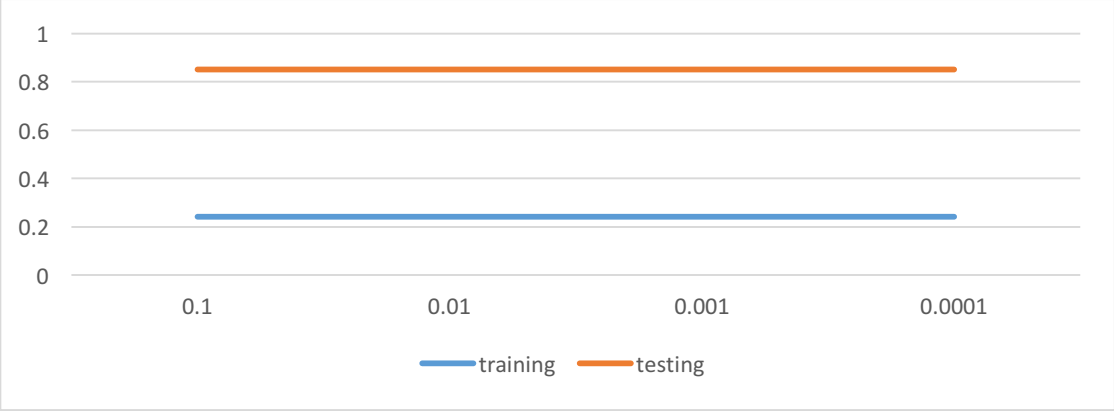
同 logistic regression，選前 6000 筆 all features 當 training data，做 logistic regression，準確率為 0.85105

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：如果沒有做 feature normalization，我的程式會在 exp 的地方 overflow。因此，得先 feature normalization，把數值壓低。

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：



由圖可知，只取前 6000 筆在 testing 表現的不錯，但在 training 方面表現不好。可能原因為前 6000 筆的 distribution 與 testing data 相似，但無法涵蓋 training data。

5.請討論你認為哪個 attribute 對結果影響最大？

我認為 age, fnlwgt, capital_gain, capital_loss, hours_per_week 這五項影響最大，但在 kaggle 上成績並不理想，可能還需考慮其他 attribute。