學號：B03901144 系級： 電機四 姓名：郭彥滕

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

答：

logistic regression的結果較佳

2.請說明你實作的best model，其訓練方式和準確率為何？

答：

運用xgboost所支持的classification方法，能成功讓準確率提升一點，但需要設定些參數，以避免overfitt的問題

還有加入一些重點feature的二次項，可以提高一些

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：

對於這次的訓練資料，需要對他做標準化，才能去做訓練，因為使用hot encoding的關係，feature大都為0或1，某些連續性的feature，他的數值比較大，就會讓整個程式壞掉，所以需要對它標準化。

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：

可以提高些微準確率

5.請討論你認為哪個attribute對結果影響最大？

就直觀而言，像是每星期工作多久應該是影響頗大的attribute

而實際上去觀察他的weight也是如此，他的係數比其他的高出一些

後來也有發現像是年齡也是一個重要的指標。