

學號：B03902084 系級：資工四 姓名：王藝霖

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

答：

logistic regression 稍佳，兩種方法皆有 feature normalization  
generative model: 0.8438, logistic model: 0.84766

**2.請說明你實作的best model，其訓練方式和準確率為何？**

答：

我的 best model 使用的方法是兩層分別為 64 和 32 的 neural network，feature 有經過normalization，loss = categorical\_crossentropy，optimizer = sgd。準確率列出兩次，第一次為在 kaggle deadline 前，另一次為上傳到 github 的 model 結果。

kaggle best: 0.858665  
upload model: 0.85694

**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: 0.8438  
not normalization: 0.78379

在沒有 normalization 的狀況下，因為有些數值相對其他數值相比非常的大，因此很容易主導整個 gradient 的方向，但那些比較大的數值並不一定是比較有用的特徵，這樣下來的結果就是整個更新的方向被值很大卻不重要的特徵主導，因此結果並不是那麼好。

**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：

no regularization: 0.8438  
regularization: 0.845155, lambda = 0.01

因為 regularization 是在 loss 中加上 weight，目的是希望讓 weight 越小越好。也能夠防止一些 overfit 的發生。在我的實驗中，有 regularization 和沒有的差別並沒有很大，但是有 regularization 的結果稍微好一些些，可能是因為沒有 regularization 的時候有一點 overfit 的問題。

**5.請討論你認為哪個attribute對結果影響最大？**

答：

age，這個 feature 影響最大。

因為在 logistic regression 得到的 weight 中，age 的絕對值最大。因此只要 age 有變動，最後的 predict 結果也會較大的反應 age 的改變。