學號：B03901149 系級： 電機四 姓名：陳咸嘉

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

答：

本題使用的 data 均已 Normalized。且皆取 10%的 training data 作訓練，剩餘的 90%拿來做 Validation set。Logistic Model 重複訓練 10000 次。

	Generative Model	Logistic Model
Accuracy	75.50%	84.73%

由上表顯示，Logistic Model 訓練的結果較佳。

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

本題使用的 data 均已 Normalized。且皆取 10%的 training data 作訓練，剩餘的 90%拿來做 Validation set。

Y_train 讀入後，每一行從一個數字變成兩個數字(分別代表是否為 class 0 和 class 1)，以利進行 2 class 的 classification。將 normalized 後的 data 以 106 維的 input layer 輸入 keras 的訓練模型，總共有 5 層 Neuron Layer，active mode 設為 sigmoid，output layer 是 2 維的 output，採用 softmax。誤差函數為 categorical_crossentropy，optimizer 設為 adam。

	Best Model (Keras)
Accuracy	85.18%

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：

下列分別以 10%的 training data 作訓練，90%作為 validation set，測量準確率。

表中的數字均代表 90%的 validation set 測量出的準確率

Model Normalized	Generative Model	Logistic Model	Best Model (Keras)
No	75.60%	84.73%	85.18%
Yes	75.50%	79.78%	75.80%

由上表顯示，除了 Generative Model 幾乎不變之外(因為反矩陣不存在，僅能計算 pseudo inverse matrix)。其餘訓練模型皆以有 Normalization 的較佳，因為年齡、工時、薪水等等變數儲存的方式是以實際數字(數十到數百萬)，和其他變數有顯著的差距(僅有 0、1)

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：

設 learning rate * lambda = 0.00001

Regularization	Logistic Model
No	84.73%
Yes	77.18%

當加入正規化項時，準確率反而變得更低了。

5.請討論你認為哪個 attribute 對結果影響最大？

Normalization 的影響遠大於 Regularization，正規化只是讓函數更加平滑，並無法縮小各變數數量級差距，因為變數數量級大的(如年齡、工時)容易主導該項的係數學習成效，而忽略數量級小的影響力(如國家等只有 0、1 的選項)。