

學號：R06922128 系級：資工碩一 姓名：楊碩礪

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

generative model accuracy: training set=0.7591 testing set= $(0.76231+0.76523)/2=0.76377$

logistic regression accuracy: training set=0.8526 testing set= $(0.85124+0.85380)/2=0.85252$

logistic regression的準確率較高。

2.請說明你實作的best model，其訓練方式和準確率為何？

答：使用neural network，網路架構hidden layer的neuron個數為64 16，輸出層為2維做softmax。gradient descent用adam，learning rate=0.01，L2 regularizer=0.01。

accuracy: training set=0.8659 testing set= $(0.85063+0.85995)/2=0.85529$

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：在logistic regression中，沒有做normalization的話準確率大概只有0.80，但是理論上normalization並不會對logistic regression的結果產生影響，原因是feature中有一欄的scale遠大於其他feature，讓error surface變得很橢圓(高維空間上)，造成gradient descent要走到最低點時間過長。

在generative model中則無影響。

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：logistic regression 加了regularization後(從0.0001~1)，正確率都在0.85~0.854之間跳動，沒有太大變化，推測可能是training set跟testing set的分佈接近一樣，所造成的結果。

5.請討論你認為哪個attribute對結果影響最大？

從logistic regression 的結果看來，capital\_gain這一欄位的參數最大(有做normalization)。可以合理的推測，一個人的資本利得對年收入的影響不僅是正相關，也是最要的一個變因。