學號：R07943095 系級：電子所碩一 姓名：劉世棠

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

結果與老師影片中所說的相同，logistic regression較generative model有更高的準確率，以下為實驗結果：

|  |  |  |
| --- | --- | --- |
|  | Public set | Private set |
| Logistic regression | 86.031% | 85.640% |
| Generative model | 84.643% | 84.105% |

1. 請說明你實作的best model，其訓練方式和準確率為何？

Best model 是使用logistic regression和adagrad並取0.5次方、1次方、1.5次方，其準確率在public set 上達到86.031%

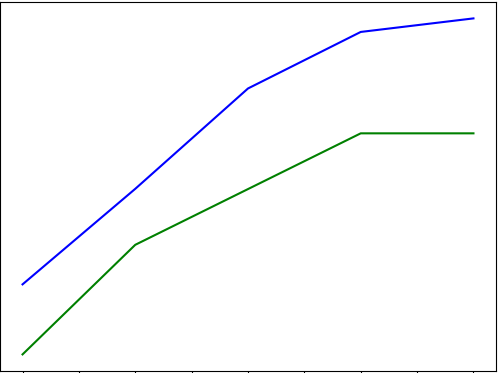
1. 請實作輸入特徵標準化(feature normalization)並討論其對於你的模型準確率的影響

在尚未實作feature normalization時準確率一直上不去（最高79%），實作後可以來到86%，可見在logistic regression上有顯著的影響。

相對於logistic regression，generative model在尚未做normalization時已可以達到84%的準確率，我想這是因為generative model的過程已有考慮平均了，故再做標準化其實沒有太大的改變。

1. 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。

我實驗的方法為使用sklearn.model\_selection的train\_test\_split將training set 分成兩份，並對不同的正規化參數做實驗，隨著正規化參數變大，整體精確度下降，而testing set和training set之間的差距有變小，但是因為整體精確度降低，故不實作。



**藍色為training loss**

**綠色為testing loss**  
  
討論：此圖的橫軸從左到右為lamda = 0.1、0.01、0.001、0，縱軸為準確度，可知lamda小則loss小，lamda 大則train loss和testing loss的gap小。

1. 請討論你認為哪個attribute 對結果影響最大？

經實驗發現weight最高的為第212項（0.5次方的age），故我認為age影響最大。