

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% |

2. 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%

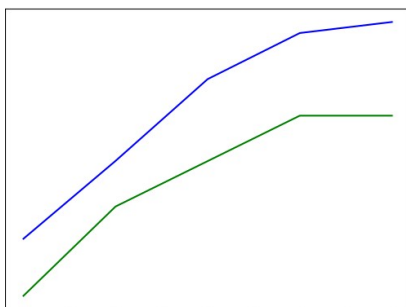
3. 請實作輸入特徵標準化(feature normalization)並討論其對於你的模型準確率的影響

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。

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。

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 $\lambda = 0.1$ 、 0.01 、 0.001 、 0 , 縱軸為準確度, 可知 λ 小則 loss 小, λ 大則 train loss 和 testing loss 的 gap 小。

5. 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。