

學號：R06921066 系級：電機碩一 姓名：劉宇閔

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

答：

使用 logistic regression 時準確率平均可以達到 0.85，但是使用 generative 時只有 0.79。

有可能是因為數據 imbalance 且數據量相差過大所造成的結果

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

best model 使用的是二次的 w 且去除一些 featurec 後進行 min.max 標準化

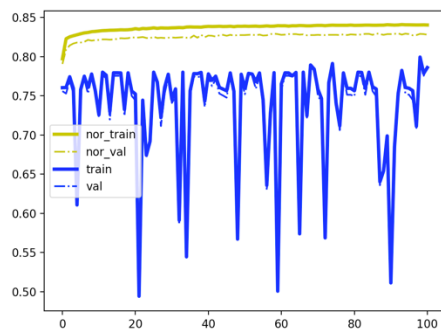
數據分割為 batch size=256 並使用 adam 演算法來進行修正，再進行了 3 萬個 epoch 後丟上 kaggle 可以得到 private 為 0.85517

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：

此數據為迭代 100 次的結果，如果做高次迭代後亦會得到類似的圖形，

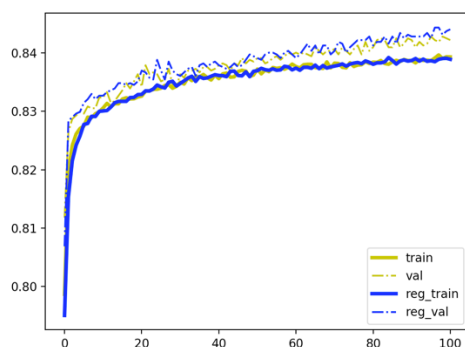
可以發現沒有做標準化時準確率會有大幅的震盪且不易收斂。



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：

使用 regularization 後 model 收斂較佳，但經過實驗後的準確率並沒有顯著的上升，推測是此種做法的模型已達全域最佳解。



5.請討論你認為哪個 attribute 對結果影響最大？

以結果來說 capital gain 影響算是最劇烈的，且在 train 完後的 model 中也可以發現他的權重植最高。