

1.請說明你實作的 **generative model**，其訓練方式和準確率為何？

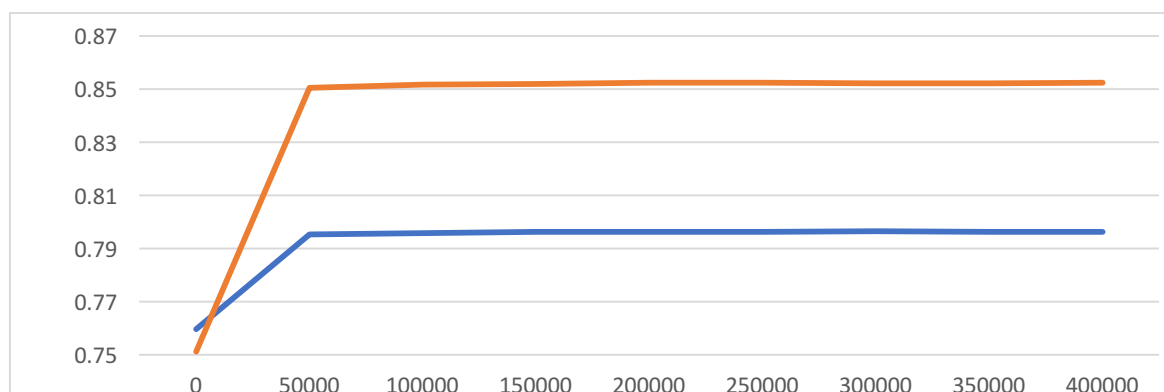
- 1) 以 One-hot encoding 的 discrete 資料與 continuous 資料作為訓練資料。
- 2) 調整 feature 數入與否或是否需要進行 feature scaling。
- 3) 以 random (輸入 seed 以確保再現性) 挑選大約 20%的資料作為 validation set，其餘作為 training set。
- 4) 將所有 training data 輸入 generative model，產生 sigmoid function 所需之變數。
- 5) 由 sigmoid function 得出的結果若大於 0.5 則視為結果為 1。
- 6) 準確率較 discriminative model 低，準確率大約為 76%。

2.請說明你實作的 **discriminative model**，其訓練方式和準確率為何？

- 1) 以 One-hot encoding 的 discrete 資料與 continuous 資料作為訓練資料。
- 2) 篩選 features, feature normalization 或進行 feature modification (加入其二次方項以提升變異程度)。
- 3) 以 random (輸入 seed 以確保再現性) 挑選大約 20%的資料作為 validation set，其餘作為 training set。
- 4) 在 logistic regression 中，對於參數訓練使用由 maximum likelihood (使用 sigmoid function 計算機率) 得出的 cross entropy 得出 gradient descent，並加入 adagrad 與 regularization 等 modification method。
- 5) 調整訓練時的 batch, epoch 數量與 learning rate, regularization lambda。
- 6) Model 的驗證中採 sigmoid function 得出的結果與 Y_train 的差值小於 0.5 即屬於預測正確。
- 7) 準確率較 generative model 高，若將 continuous feature 做 scaling，經由訓練可輕易達到 85%以上的準確率。

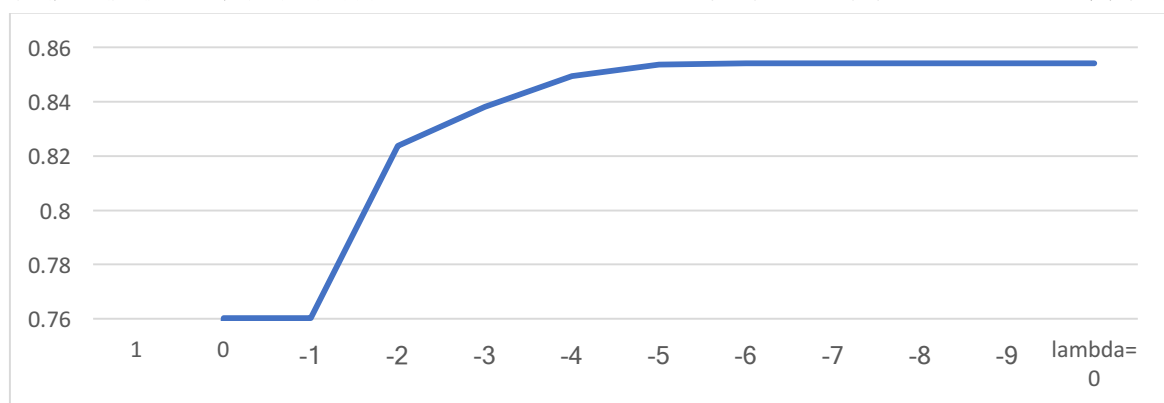
3.請實作輸入特徵標準化(**feature normalization**)，並討論其對於你的模型準確率的影響。

在此資料下針對所有 continuous attribute 進行 normalization。結果顯示 normalization 具有使預測準確率上升的效果，由於觀察 raw data 會發現各筆資料在 continuous attribute 中的變異程度極大，可能會為 gradient descent 的方向造成雜訊，且巨量的數值很有機會導致在計算 sigmoid function 時的 overflow，又會降低準確率的收斂速度。經由此項操作後，可使 continuous 資料與 one-hot encoding 的 discrete 資料數值盡量貼近，可提升 gradient descent 的效率。(X axis: 訓練資料筆數; Y axis: 預測準確率)



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。

調整 regularization 的參數 λ 值, 以同樣的參數與 model 進行訓練的準確率比較。根據結果, 可得知 λ 值在這樣的模型下不可過高, 否則無法使參數回歸至使 cost 最低的結果, 此外 regularization 對預測準確率的提升效果有限, 在某些訓練 epoch 較大量的 model 中可能可以抑制 overfit 的結果, 但由於此 model 中的預測率收斂速度快, 導致效果有限。(X axis: λ 參數的對數值; Y axis: 預測準確率)



5. 請討論你認為哪個 **attribute** 對結果影響最大?

依據 Data 提供的 attribute, 在刪除各個 attribute 其中的 feature 資料後, 以同樣的參數與 model 進行訓練的準確率比較。根據結果, 在刪除 capital gain 後對預測率影響較大, education 影響則為其次, 因此是否能夠 >50K 與這幾項 attribute 叫相關, 在直覺上也同樣合理。(X axis: 投入僅刪除該項 attribute 後的 training data 資料; Y axis: 預測準確率)。

