學號：T10902303系級： 計資一  姓名：梁高浚

1. (1%) 請比較說明generative model、logistic regression兩者的異同為何？再分別列出本次使用的資料中五個分得正確/不正確的sample，並說明為什麼如此？

Logistic regression是discriminative的方法，而generative model是generative的方法。在logistic regression中, 我們可以用gradient descent去迭代，算出optimal w and b, 使到loss function達到minimum. 而generative model的話，就是先算u1, u2, 和Σ ,然後算出w和b.

兩者算出的w和b是不一樣的。

五個分的正確的sample:

1. i:1028, y:0, yhat: 0.00670425

2. i:1029, y:0, yhat: 0.28461435

3. i:1030, y:0, yhat: 0.00968792

4. i:1031, y:0, yhat: 0.02517578

5. i:1032, y:0, yhat: 0.23454912

五個分得不正確的sample:

1. i:3764, y:0, yhat: 0.7456637

2. i:3762, y:1, yhat: 0.28589513

3. i:3753, y:1, yhat: 0.44111478

4. i:3749, y:1: yhat: 0.15594321

5. i:3743, y:0, yhat: 0.66042516

他們在logistic regression下是分的不正確的，原因可能是logistic regression這個model的complexity不足，沒辦法capture到feature之間的重要性。而且，logistic regression需要對data進行假設，他假設了feature之間是independent的，但這個在實際情況未必會發生。就比如在這幾個sample裡面，有一些feature可能中間有關聯，例如是never-married和age。最後logistic的準確率只有大概84%,過不了strong baseline. 還有就是看了一下其他的sample, 發現當y=1的時候，準確率不足，所以可能是最後算出的weighting和bias偏向y=0的sample

2. (1%) 請實作兩種feature scaling的方法 (feature normalization, feature standardization)，並說明哪種方法適合用在本次作業？

如果使用normalization, accuracy是0.8483979936533934

如果使用standardization, accuracy是0.8288463507011977

normalization把value會scale到[0,1]區間裡面，而standardization則是把data rescale至mean 0, std = 1的normal distribution裡面。從上述結果可見，本次作業適合使用normalization,而standardization的accuracy比較低，可能是因為data本身並不服從normal distribution, 應該用normalization方法比較好。

3. (1%) 請說明你實作的best model及其背後「原理」為何？你覺得這次作業的dataset比較適合哪個model？為什麼？

我的best model還是使用了logistic regression，準確率只有84%，過不了strong baseline。這個model的原理就是使用sigmoid function去預測最後P(Y=0)和P(Y=1)的概率，並且使用gradient descent去更新每次的w和b, 最後使用w和b給出預測的概率。我最後使用了grid search去尋找最佳的threshold，發現最佳threshold在53%左右，並不是50%。 不過，這個是在validation set的結果，有可能會導致overfit, 因此在最後提交的result並沒有超過private leaderboard的strong baseline。我覺得這次作業的Dataset比較適合一些deep learning model，比如neural network。因為feature有很多，需要一個比較complex的model去generalize data中間的規律，而logistic regression相比還是稍微簡單，沒辦法超越85%以上的準確率。