

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

答：訓練的 feature 是助教給的 feature 中拿掉?\_workclass、?\_occupatio 和?\_native\_country，因為我認為這些不確定的資料對訓練模型並不會往好的方向，甚至是往錯誤的方向前進。用全部的資料算出一高斯分佈，在 kaggle 得到了 0.84621 的準確率。

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

答：訓練拿的 feature 同問題一，在訓練資料中超過 50k 和低於 50k 的數量分別是 7840 和 24721 筆，為了避免模型偏向某個預測結果，故在一個 batch 中，會分別隨機取 7841 筆資料，總共 15682 筆的資料，iteration 二千次並實作了 adagrad。在 kaggle 得到了 0.845 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。

答：

模型準確率在 kaggle 上會變差。

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：

我做了 lamda=1的模型，模型的正確率在kaggle上下降了0.04。

5.請討論你認為哪個 attribute 對結果影響最大？

答：

訓練完每個 feature 的 weight 都差不多，並沒有一個特別大的 weight，認為每個 feature 對預測結果影響都差不多。