

學號：R07922104 系級：資工碩一 姓名：林傳祐

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

generative model : 0.84117

logistic regression : 0.85284

我的 logistic regression 得到比較高的準確率

2. 請說明你實作的 best model，其訓練方式和準確率為何？

我使用的是 SVM，獲得 0.86414 的準確率

除了 normalization 外，還使用 scikit-learn 的 VarianceThreshold()

可以把一些 feature 很多都是 0 or 1 的丟掉

做起來比使用所有 feature 還要好

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

若沒有使用 normalization，基本上是 train 不起來的，因為不同 feature 間的數值範圍差距過大。

我觀察 train 的 loss 是不停的上下震盪(0.2~0.7)，完全無法下降

準確率也是非常慘的 : 0.79105

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。

Lamda	0.1	1	10
acc	0.84092	0.84043	0.83884

似乎沒有什麼影響，但太大的時候，反而也是會變得做的不好

仔細觀察的話，w 似乎都不是太大，所以沒有很大的差異

5. 請討論你認為哪個 attribute 對結果影響最大？

我認為 fnlwgt 對結果影響最大，一方面他的數值範圍很廣，對每個人的意義都不相同，而且如果沒有好好處理它，結果的準確性也會有很大的變化