

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

Logistic regression 較佳，準確率為 private：0.85308，public：0.85577。

而 generative model 的準確率只有 private：0.80186，public：0.81142。

2. 請說明你實作的 best model，其訓練方式和準確率為何？

Best model 使用的是 logistic regression，準確率為 0.85577。

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 generative model 中，不使用標準化的準確率為 private：0.77238，public：0.76633。而使用標準化的準確率為 private：0.80186，public：0.81142

在 logistic regression 中，我的程式在不使用標準化的狀況下無法正常跑出結果。而使用標準化的準確率為 private：0.85308，public：0.85577。

由這兩個結果可知，使用標準化後的準確率有明顯的提升。

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。

在 logistic regression 中，再跑較少次的狀況下，不使用正規化的準確率為 private：0.84768，public：0.84877。

使用正規化的準確率為 private：0.84510，public：0.84668。

應該是我選的 lamda 不好，導致準確率稍微下降。

5. 請討論你認為哪個 attribute 對結果影響最大？

Capital gain，因為 train 完之後他的權重最大。