

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

	Public score	Private score	average
generative model	0.84643	0.84117	0.8438
logistic regression	0.84778	0.84682	0.8473

logistic regression 準確率較佳。

2. 請說明你實作的 best model，其訓練方式和準確率為何？

首先，將 fnlwgt 此項去除，因為在統計資料中代表此項目的「人數」，和此人的薪水沒有直接相關。也嘗試過把 fnlwgt 在 gradient 中作加權，但在 validation set 的準確率比直接去除差。接下來用二次式去 fit 每個 feature，由於大部分數字是 0 或 1，二次式和一次式是一樣的，所以沒有這麼容易 overfit，在 validation set 表現也是最好的。

Public score	Private score	average
0.85638	0.85566	0.85602

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

Public score	Private score	average
0.85208	0.85124	0.85166

平均的準確率增加了 0.00436。

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 validation set 中以 $\lambda=1e-5$ 的表現最好，故以此作代表。

Public score	Private score	average
0.85343	0.85087	0.85215

平均的準確率增加了 0.00485。

5. 請討論你認為哪個 attribute 對結果影響最大？

regularization 的影響較大，但兩者影響的幅度差不多。