

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

答：

generative model private:0.84215 public:0.84533

logistic model private:0.85333 public:0.85773

logistic model 較佳

2.請說明你實作的best model，其訓練方式和準確率為何？

答：

我採用logistic model，然後feature採用每筆資料的一次方、還有第1,4筆資料的二次方，最後再用regulation =30去train。

我的準確率是：private:0.85333 public:0.85773

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：

with normalization private:0.85333 public:0.85773

without normalization 沒有做normalization的feature會overflow train 不出來

有normalization 的才有結果

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：

with regularization private:0.85333 public:0.85773

without regularization private:0.85517 public:0.85823

我取最後800筆資料做validation時，with regulation的模型有比較好的正確率，可是在kaggle上卻沒有同樣的效果。

5.請討論你認為哪個attribute對結果影響最大？

我認為影響最大的attribute是第1,4筆資料，也就是age和capital_gain

因為我先做一次全部一次項的logistic model後，發現這兩個attribute的weight最大