

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

答：

Generative Model: $\text{avg}(0.80982,0.80235)=0.806085$

Logistic Regression: $\text{avg}(0.85859,0.85554)=0.857065$

Logistic Regression 較佳

2.請說明你實作的 **best model**，其訓練方式和準確率為何？

答：

使用 **Logistic Regression**，取所有一次項及前六

Feature(age,fnlwgt,sex,capital_gain,capital_loss,hours_per_week)二次項做 **Normalization** 後，以 **Adagrad** 進行 **Gradient Decent**。

Public Score:0.85859

Private Score:0.85554

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：

取所有一次項及前六

Feature(age,fnlwgt,sex,capital_gain,capital_loss,hours_per_week)二次項，使用 **Adagrad** 進行 **Gradient Decent**，**Iteration** 10 萬次，結果如下。

無 **Normalization**: $\text{avg}(0.80420,0.80945)=0.806825$

有 **Normalization**: $\text{avg}(0.85859,0.85554)=0.857065$

經過 **Normalization** 後 **Learning Rate** 可以設至 1，而無 **Normalization** 只能設至 0.0001。

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：

Lambda	Public Score	Private Score	Average Score
0	0.85859	0.85554	0.857065
0.01	0.85859	0.85554	0.857065
0.1	0.85859	0.85542	0.857005

Regularization 並沒有造成太大的影響，而當 **Lambda** 大至 0.1 時，準確率反而下降。

5.請討論你認為哪個 **attribute** 對結果影響最大？

抽取單一 **Attribute** 做 Logistic Regression，結果如下：

Feature	Public Score	Private Score
age	0.75208	0.74524
workclass	0.76891	0.76513
fnlwgt	0.76523	0.76231
education	0.78009	0.78000
education_num	0.76523	0.76231
marital_status	0.76523	0.76231
occupation	0.76523	0.76231
relationship	0.76523	0.76231
race	0.76523	0.76231
sex	0.76523	0.76231
capital_gain	0.80331	0.79928
capital_loss	0.77407	0.77042
hours_per_week	0.75638	0.75506
native country	0.76523	0.76231

其中 **Capital_gain** 的準確率可以達到八成，我認為對結果影響最大。