

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

答：

1) Accuracy (kaggle 的準確度)

	Pulic	Private	平均
Generative model	0.82653	0.82520	0.825865
Logistic regression	0.85466	0.85112	0.85289

2) 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：

1) 訓練方式

使用 logistic regression，optimizer 使用 adagrad，並且做 early stopping 的技巧防止 overfitting。features 則是挑選助教給處理過的，並未特別挑選，除原本的 features 外，並將連續的參數做平方加入當新的 features。

2) Accuracy

Pulic	Private	平均
0.85663	0.85321	0.85492

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：

1) Accuracy (kaggle 的準確度)

	Pulic	Private	平均
Normalization	0.85466	0.85112	0.85289
Without Normalization	0.79717	0.79437	0.79577

2) 討論

沒有 Normalization 的準確度明顯低很多，且在訓練的過程中，validation set的 accuracy 相當浮動，且整體訓練速度較慢。

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：

1) Accuracy (kaggle 的準確度)

	Pulic	Private	平均
$\lambda = 0.1$	0.80540	0.80481	0.805105
$\lambda = 0.01$	0.82886	0.82545	0.827155
$\lambda = 0.001$	0.83501	0.83245	0.83373
$\lambda = 0.0001$	0.84778	0.84449	0.846135

2) 討論

可能是因為有做 early stopping 的關係，可以發現 regularization 用在自身的 model 沒有幫助，甚至 λ 太大，還會出現 underfitting 的現象。

5.請討論你認為哪個attribute對結果影響最大？

答：

Capital gain，原因則是因為在已經經過 normalization 後，其權重相當的大，遠大於其他的權重，雖然 Capital gain 大多數人為 0，但顯然這仍然是個相當重要的指標。