

學號：R05943136 系級：電子所碩二 姓名：盧真玄

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

答：

Model	Accuracy(public)	Accuracy(private)
Logistic regression model	0.85393	0.85124
Generative model	0.84570	0.84203

結果 Logistic regression 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：

我的 best model 為 Gradient Boosting，其方法是對一份數據建立多個模型，再由這些模型的好壞決定權重，因此得到更好的預測結果。

我的 kaggle 正確率分別為 public: 0.87506 private:0.87188

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：

我針對 Logistic model 作實驗結果如下

Normalization	Accuracy(public)	Accuracy(private)
Yes	0.85393	0.85124
No	0.65773	0.65532

結果顯示在相同 feature 底下，有做 normalize 正確率會大幅提升。

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。

答：

我針對 Logistic model 作實驗結果如下

Regularization	Accuracy(public)	Accuracy(private)
Yes	0.85393	0.85112
No	0.85393	0.85124

結果加了 regularization 後，對模型準確率的影響並不大，我猜是 feature 的選擇，使得模型並沒有太過複雜，因此沒有產生 overfitting。

5.請討論你認為哪個 attribute 對結果影響最大？

答：

下表為使用 sklearn tool 內建好的 function，可直接跑出各 feature 的重要性，藍色長度代表重要性的百分比，由此圖可知 Doctorate 的影響最大。

