

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

答：

取 106 個 feature

generative model 準確度 0.84

logistic regression 準確度 0.79

generative 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：

用 sklearn 裡面的 GradientBoostingClassifier 套件 得到 public 0.87 的準確度

原理:三個臭皮匠>一個諸葛亮

Gradient Boost 是一個框架，里面可以套入很多不同的算法。

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：

未做 feature scaling 前，得到準確度 0，做完 normalization 後得到準確度 0.84，根據我的觀察，是因為在找 inverse matrix 的時候，因為懸殊的 scale 產生巨大的誤差，因此造成不精準。

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：

未做 feature scaling 前，得到的準確度為 0.79，做完 normalization 後得到準確度亦為 0.79，影響不大。

5.請討論你認為哪個 attribute 對結果影響最大？

利用 sklearn 裡的套件作圖後得到第五個 feature 也就是 capital\_loss 的 importance 最大