

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

答：

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：

以keras實作類神經網路(一層隱藏層106個神經元relu作為activation function)

adam做優化，batch_size=300, epoch = 150

在kaggle上public得到0.86093的正確率

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：

沒有使用特徵標準化會使其不太能train起來，得到的結果遠不如有做特徵標準化的結果，可能是因為卡在某些local min，使其訓練結果往往得不到好的結果。

而在logistic regression時若沒有用feature normalization的話會使其訓練速度極度緩慢

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：

做了正規化之後，雖然在訓練集上的正確率有些微下降，但是在調整正規化的大小之後，可以發現在testing_set上面的結果變好很多，可能是沒有做正規化時有overfitting的情況發生，而正規化之後使得參數變的平滑，讓overfitting的情況消失了。

5.請討論你認為哪個attribute對結果影響最大？

我認為是capital_loss。

以同樣的方法、模型去做training(同樣的複雜度，同樣的訓練次數、batch_size)

若我以單個變數做training，以capital_loss可以在testing_set上得到最好的結果。

若我以105個變數做training，少掉capital_loss，使其在testing_set上得到最差的表現。

另外以人腦來想這也是蠻合理的，因為如果是沒錢的人，資本的損失本來就應該比有錢人少，因此我認為是capital_loss