

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

generative model 分數：public 分數是 0.76523、private 分數是 0.76231

logistic regression 分數：public 分數是 0.84692、private 分數是 0.84461

從 kaggle 分數上來看，可以明顯看出用 logistic 的方式做出來的分數比 generative 高很多，這是因為 generative 是線性的方式來把資料作分類，而 logistic 有經過 sigmoid function 所以可以處理非線性的問題，因此 logistic 的分數會高出許多。

**2.請說明你實作的 best model，其訓練方式和準確率為何？**

我的 model 分數：public 分數是 0.86056、private 分數是 0.85345

我的 best model 使用的是 Keras 的套件來做深度學習，在 train 資料之前有做 normalization，模式使用了 106 個神經元，activation 是使用了 relu 跟 sigmoid，同時也使用 dropout 減少模式會 overfitting 的機率，最後 loss function 用 binary\_crossentropy，使用 adam 方式來更新 learning rate，batch 跟 epochs 的大小都給 100。

在訓練的時候適了很多方法，有給好幾層的 layer，但是效果並沒有變好，說明並不是 layer 越多，training 的效果越好。除此之外，有做 validation 的話會讓模式 overfitting 的機率下降。

**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。**

在沒有做 normalization 的時候 kaggle public 分數為 0.79，而有 normalization 則會提高到 0.84。在把 data 標準化後會讓資料變得比較好 train，在 gradient descend 的時候比較能快點找到 local minimum，所以在同樣的迭代次數下會有較佳的結果。

**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。**

我利用 logistic model 做 regularization，lamda 給 0.1 發現原本不做 regularization 時 training data 得到的分數是 0.78，但是給了 lamda 之後卻下降到 0.759，而在 testing data 上也變差了(從 0.79 掉到 0.76)。把 lamda 換成是 0.001 在 training data 得到的是 0.775021，在 testing data 上得到的是 0.779，同樣跟最初比較是比較低的。我覺得在沒有 overfitting 產生的情況下使用 regularization 好像反而會讓結果變差，因為做正規化會使得到的線變平滑，但是對點的 fitting 結果就會變差。而這件事在有 overfitting 產生的時候做正規化會讓 testing data 結果變好，但如果今天沒有 overfitting 情況的話反而就變差了。

**5.請討論你認為哪個 attribute 對結果影響最大？**

我認為是 capital loss。

我有把每個參數挑出來單獨做 training，得到的結果在 capital loss 的時候會最好，因此我認為 capital loss 對結果影響最大。