

學號：B04901146 系級： 電機三 姓名：黃禹傑

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

答：

	Training accuracy	Testing Accuracy
Generative model	0.84273	0.84299
Logistic model	0.79221	0.79264

Generative model 得到的結果較佳，高了大約 5%，然而等等可以發現在做過 feature normalization 之後，logistic model 的準確率便可以提高至相同水準。

2. 請說明你實作的 best model，其訓練方式和準確率為何？

答：

Best model 是引用 keras library 實作，讀入 training data 後先進行 normalize，接著 traing model 的架構為 input layer (106 neurons) -> one hidden layer (1024 neurons) -> output layer (1 neuron)，其中有加入 dropout (rate = 0.2)，activation function 採用 relu，output 時再經過 sigmoid function，並以以 0.5 為分界決定預測結果為 0 或 1。

Training accuracy = 0.85810

Testing accuracy = 0.84841

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：

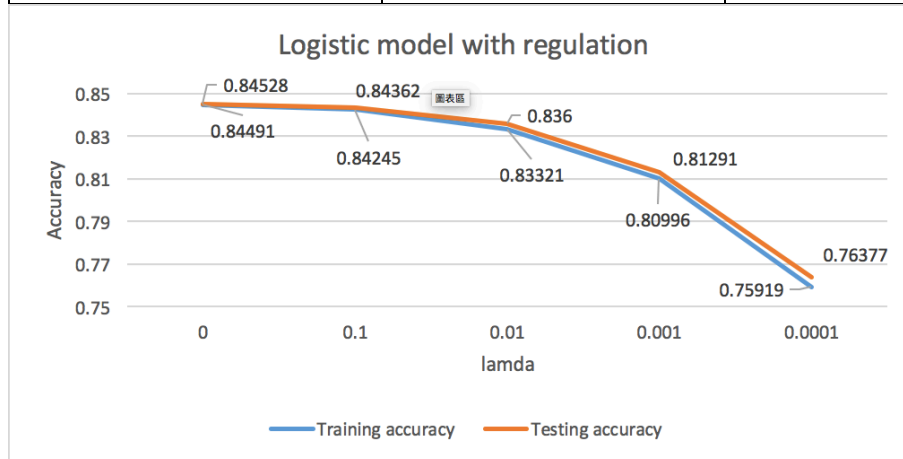
	Training accuracy	Testing accuracy
Generative with normalization	0.84230	0.84380
Logistic with normalization	0.84491	0.84528

Generative model 的準確率並沒有提高，然而 logistic model 的準確率是提高並甚至高於 generative model，可見 feature normalization 對於 logistic model 的 training 是十分有效的，而 generative model 因為是利用數學式直接推導 w 與 b，所以對於準確率的影響就小許多。

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：(with normalization)

Logistic with regulation	Training accuracy	Testing accuracy
$\lambda = 0$ (no regulation)	0.84491	0.84528
$\lambda = 0.1$	0.75919	0.76377
$\lambda = 0.01$	0.80996	0.81291
$\lambda = 0.001$	0.83321	0.83600
$\lambda = 0.0001$	0.84245	0.84362



5. 請討論你認為哪個 attribute 對結果影響最大？

在 generative model with feature normalization 下，將參數 w 以及 b 輸出以便觀察，可以發現最大的參數 w 對應到的 attribute 是 capital gain，代表 capital gain 些微的差距可以對結果造成最大的影響，另一方面而言，資本收益也確實為和一人的財力頗有關係，所以我認為 capital gain 對結果影響最大。