

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

答：

|                     | training acc. | testing public acc. | testing private acc. |
|---------------------|---------------|---------------------|----------------------|
| generative model    | 0.844595      | 0.84508             | 0.84227              |
| logistic regression | 0.808872      | 0.79815             | 0.79019              |

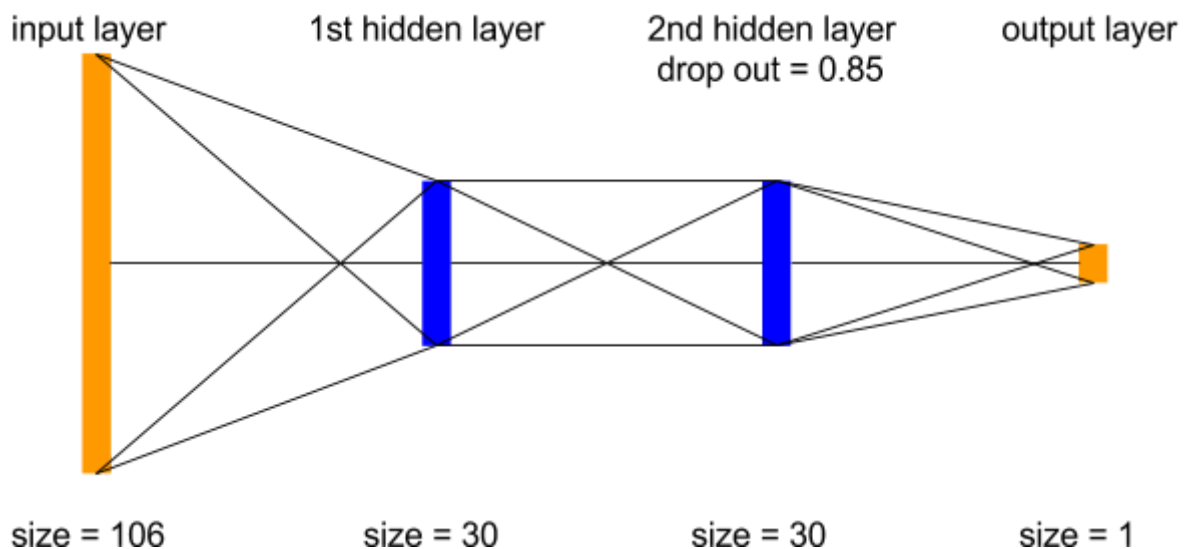
結果上來說是generative model 的準確率比較高。

推測原因是，training data 總共只有32561 筆，其實並沒有很充足。因此，generative model 對於機率分佈的額外假設可以幫助提升預測的準確率。

2.請說明你實作的best model，其訓練方式和準確率為何？

答：

我使用keras 套件來實作我的best model，下圖是我的model 架構。



我在中間加了兩層hidden layer，size = 30，並且第二層hidden layer 的drop out rate = 0.85 (用以防止over fitting)。

將batch size 取為128，使用keras 的'binary\_crossentropy' 計算我1一維output 的 loss function，總共train 了100 個epoch，結果如下表。

|          | testing public acc. | testing private acc. |
|----------|---------------------|----------------------|
| NN model | 0.85933             | 0.84977              |

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：

|                    | testing public acc. | testing private acc. |
|--------------------|---------------------|----------------------|
| w/o normalization  | 0.79619             | 0.79290              |
| with normalization | 0.79815             | 0.79019              |

如上表所示，加入normalization 對準確度並沒有影響。Normalization 的作用只是幫助training process 的收斂，因此若是model 已經收斂，normalization 就不會影像預測的準確率。

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：

|                     | testing public acc. | testing private acc. |
|---------------------|---------------------|----------------------|
| w/o regularization  | 0.81228             | 0.81341              |
| with regularization | 0.79815             | 0.79019              |

加入regulariziton 後，model 預測的準確率提高了。推測原因是加入 regularization 後對複雜的model 進行了懲罰，使得原本overfitting 的情況可以緩解。

5.請討論你認為哪個attribute對結果影響最大？

從結果來看，regularization 的影響應該最大。

由於使用logistic regression 的model 過於複雜，容易導致overfitting，因此這次可以得到兩個解決方法。一是，對model 的複雜程度進行punishment，例如加入 regularization。使得原本logistic tegression model 可以走到一個相對簡單的model。二是，一開始對model 的假設加入限制，例如generative model。在一個比較間單的function set 中去找最小誤差值。