

1. (2%) 請比較實作的 generative model 及 logistic regression 的準確率，何者較佳？請解釋為何有這種情況？

logistic regression 的結果較佳，因為 logistic regression 的結果會比較依賴 training data，而 generative model 則是會比較依賴它預測，因此在 training data 量足夠的時候，logistic regression 的結果通常會比 generative model 的預測結果來的要好。public 上 logistic 可達 0.89，generative 只有 0.88。

2. (2%) 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。接著嘗試對正規項使用不同的權重 (λ)，並討論其影響。(有關 regularization 請參考 <https://goo.gl/SSWGhf> p.35)

大概嘗試了 $\lambda = 0.0001$ 、 $\lambda = 0.0005$ ，在 validation 的結果都和 $\lambda = 0$ (未使用 regularization) 差不多。

3. (1%) 請說明你實作的 best model，其訓練方式和準確率為何？

我實做的 best model 使用了 logistic regression，training data 為助教提供已經整理過的 X_{train} 、 Y_{train} ，並且一開始先去除一些 feature，像是幾乎為 0 或是出現許多問號 (?) 的 feature，max iteration 設為 20、batch size 設為 10、learning rate 設為 0.1，最後在 kaggle 的 public 準確率為 0.89059。

4. (1%) 請實作輸入特徵標準化 (feature normalization)，並比較是否應用此技巧，會對於你的模型有何影響。

在 validation 的表現，有使用 feature normalization 的結果較佳，最後也是以有做 feature normalization 通過 strong baseline。