

HW2

1. 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？

以 generative model 所預測的準確率為 0.88067，而已 logistic regression 所預測的準確率為 0.88979，相較之下高於 generative model 所預測的結果。因 generative model 在 probability distribution 上有許多額外的假設，包含 Gaussian、Bernoulli 及 Naïve Bayes.....等假設，因此在結果預測上產生了誤差，造成準確率低於無額外假設的 logistic regression。

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 (lambda)，並討論其影響。

未使用 regularization 時，預測值的準確率為 0.88979，而在使用 regularization 後準確率出現下降的現象，在 lambda 為 0.01 時，準確率為 0.88791，而在 lambda 為 0.1 時，準確率為 0.88045，因此可以發現 regularization 在提升準確率上並沒有太大的作用，且 lambda 值越大準確率越低。

3. 請說明你實作的 best model，其訓練方式和準確率為何？

我的 best model 是利用演算法 XGBoost 來進行 training，先利用 cross-validation 找出相對較佳的參數組合，接著以這些參數組合對 training data 進行計算，最終所得的 model 在 test data 上的預測準確率可以達到 0.90507，相對於 logistic regression 有更佳的準確率。

4. 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。

有使用 normalization 時準確率為 0.88979，在沒有使用 normalization 時準確率為 0.88907，且在進行計算時會產生相當大的數值震盪，因此可以知道 normalization 對於 weight 具有提升穩定性的作用，進而提升預測準確率。