

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？

在相同的資料處理下 (only normalization) 結果圖如下

model	accuracy
Generative model	0.8749078442937187
Logistic regression	0.8763361592333211

從結果來看是 logistic regression 結果較好，原因是因為用 generative model 時，對機率分布有預先的假設(高斯分佈)，但實際上真實的資料不一定會符合這個假設，因此正確率會有一定的上限，而 logistic regression 則沒有這個限制，因此找出的結果是可能比較好的

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 (lambda)，並討論其影響。(有官 regularization 請參考 <https://goo.gl/SSWGhf> p.35)

Lambda	Training accuracy	Development accuracy
0	0.7362277288552119	0.7364541098415038
0.00001	0.741634241245136	0.7401400663472171
0.0001	0.7777800532459553	0.7762624401032068
0.001	0.7328896170387	

可以看出有用 regularization 可以讓正確率上升，實驗的結果是在 lambda 等於 0.0001 時有較好的結果，正確率上升了 4%

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

得到正確率最好的結果是使用 logistic regression，data preprocessing 部分有 normalization，並且只使用部分的 feature (134 維)。訓練的部分使用 gradient descent 加上 regularization，learning rate 為 0.1。最後在 kaggle 上得到 0.889 的正確率

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。

	Training accuracy	Development accuracy
No normalization	0.7362277288552119	0.7364541098415038
With normalization	0.8845177145197625	0.8763361592333211

實驗結果是將資料經過 normalization 後，可以讓正確率大幅度的上升，相差了 15%。