

學號：B07901014 系級：電機二 姓名：陳希格

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？

generative model accuracy : 0.8702263344146269

logistic regression Training accuracy : 0.8851730493549048

Development accuracy: 0.8765204570586067

Logistic regression 明顯較好。這是因為 generative model 已經假設了某種分布 (Gaussian Distribution)，而實際情況不一定符合該分布。

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。(有官 regularization 請參考 <https://goo.gl/SSWGhf> p.35)

λ	0	0.2	0.4	0.6	0.8	1.0
Training accuracy	0.88539	0.87898	0.87571	0.87356	0.87177	0.87063
Development accuracy	0.87762	0.87136	0.87062	0.86877	0.86601	0.86325

就結果來說，原本的 model overfitting 的情況並不嚴重，所以加上正規項並沒有表現得比較好，反而越大的權重越影響其表現。

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

基於 logistic regression 的 model，只是過濾掉一些相關係數的絕對值過小的 feature，最終 Training accuracy: 0.88546，Development accuracy = 0.88001。

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。

lambda	normalization	No normalization
Training accuracy	0.88539	0.77755
Development accuracy	0.87762	0.78013
Training loss	0.26792	3.92687
Development loss	0.28596	3.88409

沒有 feature normalization 的準確度明顯下降，loss 也較大，所以 feature normalization 可以有效幫助我們在 gradient descent 中更快的降低 loss。