

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？

比較 generative model 和 logistic regression 在 development set 和 testing data 上的準確率後，發現都是 logistic regression 的準確率較佳，generative model 與 logistic regression 不同的地方在於 generative model 會對機率分布做假設，而當資料少的時候，或許 generative model 能夠有比較好的，但當 training data 夠多的時候，generative model 的這些假設就不見得有好的效果，因此造成 logistic regression 有較佳的準確率

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)

在 epoch = 10, batch size = 8, learning rate = 0.2 的情況下，我對 logistic regression 作正規化後，發現會造成模型的 training accuracy 下降，但 development accuracy 上升，而我選了四種不同的 λ 作實驗，分別為 1, 5, 10, 20，發現 $\lambda = 5$ 或 10 對 development accuracy 的上升是有幫助的，但 $\lambda = 1$ 或 20 卻有較差的效果，其中 $\lambda = 10$ 的效果是最佳的

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

我最後選用的 model 是使用 logistic regression，然後我將 iteration 調成 1000 次, batch size = 20, learning rate = 0.005，由於在 training 1000 次的情況下，我經過多次嘗試覺得 learning rate = 0.005 是較為合適的大小，而我也挑選對 training 較有影響的前 200 個 feature 來降低 dimension，同時達到準確率提升的效果

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。

有加入特徵標準化和沒有加入特徵標準化有很大的差別，加入特徵標準化後無論是 training accuracy 或 development accuracy 都有很大的提升，應用此技巧後，對模型的影響為可以改善資料範圍不同的問題，以這次作業的 training data 為例，有許多資料的值為 0 或 1，但像 age 這項資料，範圍就有可能是 0-100，但並不代表 age 對 classification 有很大的比重，因此作 normalization 就能避免這種問題發生