

學號：B07902074

系級：資工二

姓名：張媿媿

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 的準確率較佳。

因為在資料量大的情況下 Logistic regression 是比較有優勢的，相較於 generative model 他比較不會去腦補(做一些多餘的假設)。

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

$\lambda = 0.0001$

Training accuracy: 0.8840057341798075

Development accuracy: 0.8772576483597494

$\lambda = 0.001$ :

Training accuracy(90%): 0.8837804628302273

Development accuracy(10%): 0.8772576483597494

$\lambda = 0.01$ :

Training accuracy(90%): 0.8811386442760598

Development accuracy(10%): 0.8746774788057501

$\lambda = 0.1$ :

Training accuracy(90%): 0.868113864427606

Development accuracy(10%): 0.8643568005897531

我發現有沒有做正規化似乎對我的model準確率沒有太大的影響，藉由調整

lambda大小，可以發現當lambda越大，準確率有下降的趨勢。

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

- -將age 跟 work hours per week 做 one-hot-encoding，然後用logistic regression來訓練，準確率大約有0.886。

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。

- 我發現有 normalization的話，準確率大約有88%，而若沒有的話，準確率大約有78%，可見有標準化的話，可是資料較平衡，使準確率升高