學號：R05944019 系級： 網媒一 姓名：張嘉豪

1.請說明你實作的 generative model, 其訓練方式和準確率為何？
答：用 Gaussian model 實作，一開始先將年收入大於 50k 以及小於 50k 的分成兩群，分別算出兩群的平均和 covariance matrix，算出來之後帶入最佳化 w 和 b 的公式，最後由 sigmoid(wx+b)決定預測的結果應該被歸類在大於 50k 還是小於 50k。

	Gaussian
準確率	0.84103

2.請說明你實作的 discriminative model, 其訓練方式和準確率為何？
答：10000 個 epoch 每次 batch 的數量是 80 個，loss function 是用 cross entropy，每 80 筆資料就更新一次 w 和 b 的值，一開始的時候 batch 的數量是 128，但是計算過後發現，如果使用 128 的話，每個 epoch 就會有 48 筆資料沒有使用到，因此準確率也比較低，而如果用 80 的話，每次只會有一筆資料沒有使用到。

	Batch size: 80	Batch size: 128
準確率	0.85086	0.83317

3.請實作輸入特徵標準化(feature normalization), 並討論其對於你的模型準確率的影響。
答：這次的 training data 有幾個欄位的數字很大，若沒有先將特徵標準化的話，很容易會導致 overflow，只有對‘fnlwgt’標準化的話，效果會不太好，若是將全部都標準化的話，效果也不盡理想。

	fnlwgt	age, fnlwgt, capital_gain, capital_loss, hours_per_week	全部標準化
準確率	0.75642	0.85086	0.78428

4. 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。
答：
在 $\lambda = 0.001$ 的時候，會有最好的準確率，若是將 λ 變大，準確率會先下降，到 $\lambda = 0.1$ 的時候，準確率會上升，再更大的時候準確率會下降得很快

	$\lambda = 10$	$\lambda = 1$	$\lambda = 0.1$	$\lambda = 0.01$	$\lambda = 0.001$
準確率	0.78	0. 76195	0.83270	0.80690	0.84867

5.請討論你認為哪個 attribute 對結果影響最大？

我先將大於 50k 的和小於 50k 的分成兩群，每群裡的每個 attribute 算出其平均以及標準差和中位數，用算出來的結果來分析哪個 attribute 比較能夠分辨出兩個群組。

我發現‘Married-civ-spouse’在兩群的平均相差蠻多的，兩群的平均相差在一個標準差之外，因此我認為‘Married-civ-spouse’是影響蠻大的 attribute。

	50k 以下	**50k 以上**
平均	0.33511327	0.85346257
標準差	0.47203005	0.35364419
中位數	0	1