學號：B03901103 系級： 電機三 姓名：陳學平

1.請說明你實作的generative model，其訓練方式和準確率為何？
答：
將 binary features 視為 bernoulli random variable，將 continuous features 視為 gaussian random variable。並且使用 naive-bayes classifier，也就是將各 random variable 視為 class indepenent。
對於 bernoulli random variable X，可直接從 training data 中不同 label 中 X=1 和 X=0 的數量來估計其機率分佈 P(X | Y = 1) 和 P(X | Y = 0) 。
對於 gaussian random variable Z, 則可由training data 中不同label的sample mean和sample standard deviation 作為機率模型的參數，以此得出 f(Z | Y=1) 和 f(Z | Y=0) 。
基於上述的模型，對於給定的 input X，便由 P(Y = 1 | X) / P(Y = 0 | X) 是否大於一來判別 Y = 1 或 Y = 0。
以上模型在validation set 上之準確率為 80.65%

2.請說明你實作的discriminative model，其訓練方式和準確率為何？
答：
此次實作的是logistic regression，訓練方式使用的是 gradient descent：loss function使用的是cross entropy，batch size = 32，learning rate = 0.01，iteration 次數為100次，並且每當loss function得出的值不減反增時，就把 learning rate 除以二，如此一來可以adaptive的改變learning rate。
以上模型在validation set上之準確率為 85.20 %

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：
本題針對logistic regression來回答：
原本的準確率：80.14 %
實作特徵標準化的準確率：85.20 %
實作特徵標準化後，準確率進步的原因可歸納如下：

a. gradient 會跟 X的 scale成正相關，若是X的scale大小不一，則不同的feature更新的 step size 會與各自的scale成正比，如此一來設定 learning rate的意義就被破壞
b. 從圖形上來看，normalize過後的 loss 對 features的圖形接近正圓，而沒有normalize時其圖形很可能是長橢圓，因此normalize過後做gradient descent比較容易收斂。
c. 本次的feature若沒有先normalize，則在通過sigmoid function時會發生overflow，導致數值的精確度喪失。

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：

lamda	0	0.001	0.003	0.01	0.1
accuracy	85.20%	85.23%	85.22%	85.15%	84.76%

從上表可以發現，加入 regularization 對 logistic regression 並沒有辦法在準確度上有太多的影響。觀察此次的各個feature對應到的weight，可以發現到其實並沒任何一個feature有非常大的權重，而regularization是用來限制讓權重不可太大，因此在這次的例子上，regularization並沒辦法發揮到什麼作用。

5.請討論你認為哪個attribute對結果影響最大？
capital_gain的影響最大，可由以下兩種方法驗證：
a.只選用一種feature做logistic regression，比較使用哪一個feature可以在validation set 上有較高的準確率
b.選用全部的feature作logistic regression，且feature有經過normalization。最後再比較哪一個feature對應到的weight之絕對值最大。
用以上兩種方法，皆顯示出capital_gain對結果影響最大。
直觀上的解釋：薪水高的人有比較大的機會從事資本交易，因此會對應到較高的capital_gain。