學號：B04902004 系級： 資工二 姓名：王佑安

1.請說明你實作的 generative model，其訓練方式和準確率為何？

將所有 data 標準化後丟進去照投影片的作法計算μ跟 Σ，再將算好的參數帶入 Gaussian Distribution 直接計算機率。

準確率：

Training Accuracy: 0.86029

Validation Accuracy: 0.85919

2.請說明你實作的 discriminative model，其訓練方式和準確率為何？

將所有 data 標準化後，以 cross entropy 作為 loss function 進行 gradient descent。

準確率：

Training accuracy: 0.86198

Validation accuracy: 0.85980

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。

	Generative Model		Discriminative Model	
	Training	Validation	Training	Validation
未標準化	0.84177	0.84128	0.57959	0.57919
標準化	0.86029	0.85919	0.86198	0.85980

可以看出標準化對準確度有顯著的幫助，尤其在 Discriminative Model 中，沒有標準化會讓 Gradient 浮動太大，無法有效地找到最小值。

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。

λ	0	1e-3	1
Training accuracy	0.86198	0.86198	0.86168
Validation accuracy	0.85971	0.85980	0.85949

正規化在這個例子只有極小的幫助，在 λ 過大時也會讓準確率下降

5.請討論你認為哪個 attribute 對結果影響最大？

我認為 capital gain 對結果影響最大。只選一項 attribute 當作 feature 時，capital gain 可以讓準確度達到 0.78，capital loss 則有 0.76，其他 attribute 都在 0.75 左右。只選擇 capital gain 跟 capital loss 就可以讓準確度到達 0.81。因此我認為 capital gain 是對結果影響最大的 attribute。

註：

以上所有 Validation Accuracy 是將 data 隨機選 20%作為 Validation set，測試五次平均的結果