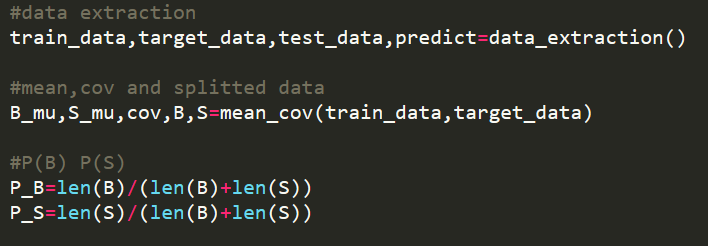
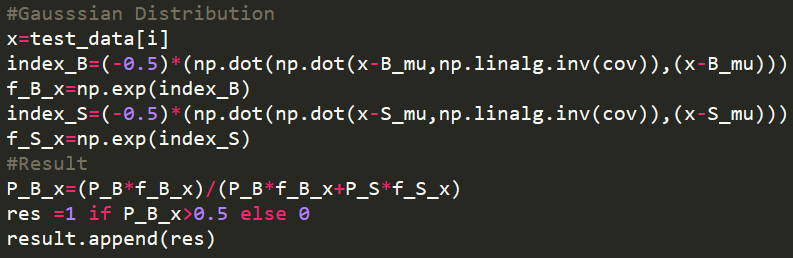
學號：R05921037 系級： 電機碩一 姓名：陳冠廷

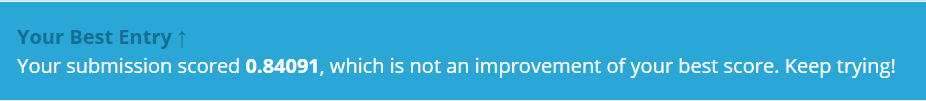
1.請說明你實作的generative model，其訓練方式和準確率為何？

答：

左圖這幾行是用來算出所需要的MEAN以及COV。

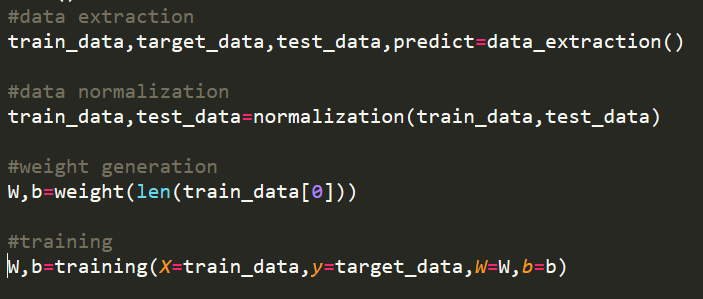
左圖則是將上面所算出來的MEAN、COV帶入高斯並求出答案。

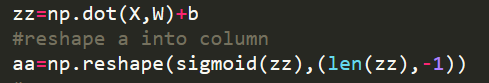
下圖則是我利用generative model 在kaggle上的結果，可以預先知道的這不會是一個最好的方法。



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

答：

如同linear model我也是先取出data，然後標準化，產生初始化參數，最後進入迴圈訓練，不一樣的是如下圖在求出zz後我需要經過一個sigmoid function。



此次的準確率就比generative model 高出了1.2%，可以看出確實discriminative model效果是比較好的。

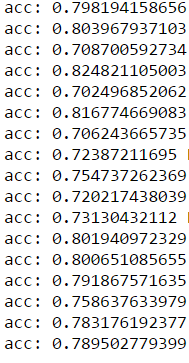
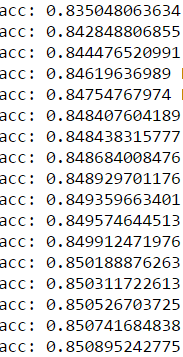


3.請實作輸入特徵標準化(feature normalization)，並討論其對於你的模型準確率的影響。

答：

下面兩張圖都是logistic regression在同樣是SGD 以及eta一樣的情況下做的比較

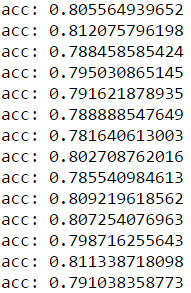
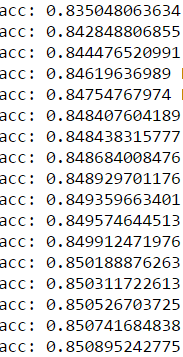
左圖是沒有標準化的數據可以看出不只反映精確度的數據亂跳而且有向下跳動的趨勢，而右圖則是有標準化的數據可以看出精確度穩定上升中。

而在generative model的部分則是差異不大，我想可能是因為generative model本身就是利用其平均以及標準差來衡量新的數據，所以如果作標準化後感覺只是把所有的資料平移縮放，對於高斯函數所界定的邊緣界線不會差太多。

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：

左圖為沒有正規化的logistic regression，其精確度一直卡在80%左右不上不下，我想主要原因是因為每一步走的步伐太大所以無法往min的地方走下去一直卡在那個凹口，而右圖則是有正規化的結果，會進入那個凹口不會卡在那邊。

5.請討論你認為哪個attribute對結果影響最大？

左圖是將106個feature對各自的1或0作平均後的圖，可以看出紅色於(33,53)也就是Married-civ-spouse以及Husband有最大值而藍色於這兩個位置剛好是反指標，所以可以確定這兩個指標較為明顯。