學號：B05901189 系級： 電機一 姓名：吳祥叡

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

(1) bernoulli:
將非 binary 的 feature 全部除以最大值,以 bernoulli distibution fit. 準確率 0.76069.
(2) gaussion: (Github 上傳這個)
用老師講義 p18 的 maximum likelihood 公式,且兩個 Class 用同一個 covariance matrix.準確率 0.84103.
(3) gaussion + bernoulli:
綜合上面兩種, binary feature 用 bernoulli, 非 binary feature 用 gaussion. 準確率 0.76290.

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

使用 Adagrad 的 learning rate.其餘跟老師講義一樣.
由 public score 評估 有 regularization 的準確率最高為 0.8553
無 regularization 的準確率為 0.85307

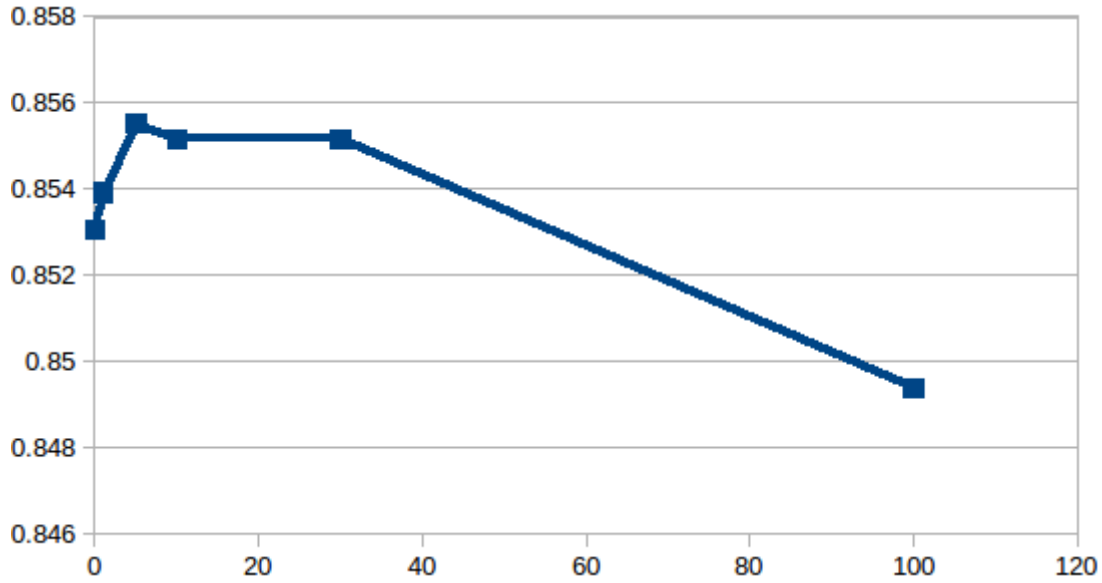
3.請實作輸入特徵標準化(feature normalization)，並討論其對於你的模型準確率的影響。
答:

用 discriminative model 比較
(1) 將所有 feature normalize: 準確率 0.82899
(2) 只將非 binary feature normalize: 準確率 0.8507
(3) 不 normalize: 0.79779
推論只將非 binary feature normalize 會有較好的節果.

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:

lambda 為 regularization 强度的參數:

λ	準確率（public score)
0	0.85307
1	0.85393
5	0.85553
10	0.85516
30	0.85516
100	0.84939



5.請討論你認為哪個 attribute 對結果影響最大?

利用 discriminative model 的參數討論各個 attribute 對結果的影響:

Attribute no.	parameter	meaning
8	-5.35773652985	Never-worked
13	-6.87991747194	Without-pay
28	-13.0546181098	Preschool
46	-4.71460491214	Priv-house-serv
78	-4.25322679631	Holand-Netherlands
91	-6.07291497807	Outlying-US(Guam-USVI-etc)

列出參數絕對值大於 3 者,因為都是負數,所以找到的都是反指標.

包括無工作經驗,學歷低於小學,荷蘭人(因為稅很重?),美國的外島...等因素.