

Machine Learning HW2

學號：R04922169 系級：資工所碩二 姓名：楊智偉

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

答：

(1) 訓練方式：

抽取助教提供的 106 維資料來使用，並且跟據薪水的高低分為兩類。用這兩類 **data** 分別算出 μ_1, μ_1 以及共同的 Σ 。藉此參數直接算得 **model** 的 w, b 再使用 **sigmoid function** 來計算 $P(C_1|x) = \sigma(w \cdot x + b)$ 。

(2) 準確率：

Public Score	Private Score
0.83931	0.84425

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

答：

(1) 訓練方式：

一樣使用 106 維資料來 **train**。利用 **Cross Entropy** 來表示 **Loss function**，接著對 w 微分之後做 **gradient descent** 來找到最佳的 w, b 。Training 過程中也使用 **Adagrad** 來實作。

(2) 準確率 (learning rate=1.0, iteration=1000)：

Public Score	Private Score
0.84533	0.84719

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：

No normalization	Feature normalization
0.79681	0.84533
0.79560	0.84719

上表比較是否使用 **feature normalization** 所得到的準確率差異。兩種 **case** 均是在 **learning rate=1.0**，**iteration=1000** 的情況下執行。很明顯看到使用 **feature normalization** 的方法明顯得到準確率的提升。

原因可能是因為前 6 維的資料中有很多資料的數值大小差異很大，導致在做 **gradient descent** 的時候沒辦法在很短的時間、很少的 **iteration(5000)**次數中得到很好的結果。而做過 **normalization** 的方法則會比較容易達到最佳解的區域。

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：

No regularization	Regularization
0.84533	0.84545
0.84719	0.84658

上表比較是否使用 regularization 所得到的準確率差異。兩種 case 均是在 learning rate=1.0，iteration=1000，使用 feature normalization 的情況下執行。這次的結果準確率就沒有太大的提升了，甚至在某些情況下準確率反而變低。

原因可能是這次 classification 問題所得出來的參數都偏小，regularization 的效果就沒有這麼明顯，可能要再將 λ 的數值調得大一些才能看出不同。但是 λ 數值過大也會有不好的影響，會讓 model 太過傾向於找到參數較小的值，反而讓準確率更加不好。

5. 請討論你認為哪個 attribute 對結果影響最大？

答：

	Public Score	Private Score
All feature	0.83649	0.84265
Age	0.83489	0.84081
fnlwt	0.83759	0.84277
Sex	0.83673	0.84228
C-gain	0.83391	0.83675
C-loss	0.83391	0.84044
Hours	0.83649	0.84093
Workclass	0.83698	0.83909
Education	0.82862	0.83528
Marry	0.83710	0.84289
Occupation	0.82985	0.83663
Relation	0.83587	0.84130
Race	0.83771	0.84339
Country	0.83771	0.84265

上表是將某一個 feature 丟掉後，使用 generative model 來測試是否有下降很多準確度。我們可以從結果看到，拋棄 Education 與 Occupation 的欄位後，準確率下降最多。所以合理的推測，Education 以及 Occupation 的影響相對來說是比較大的。