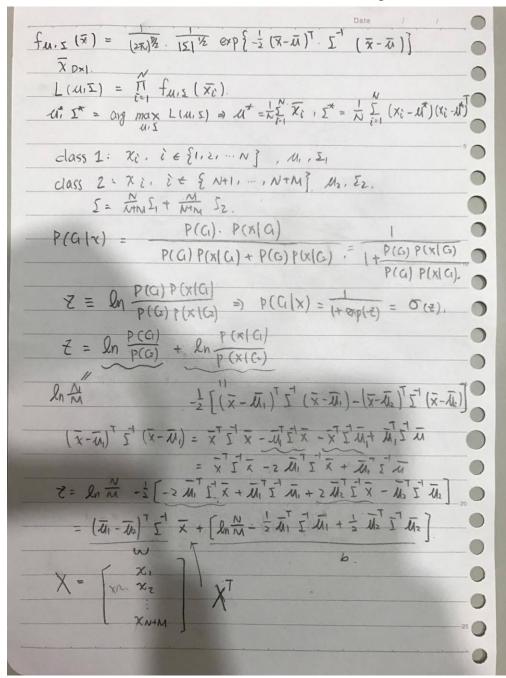
Machine Learning Hw2 – Report

學號: B03901026 系級:電機三 姓名:許凱傑

1. 請說明你實作的 generative model, 其訓練方式和準確率為何? 先將 train data 分成兩類,並分別算出 mean 跟 covariance matrix, 再得到加權 後的 covariance matrix, 如下圖計算出 w 跟 b。準確率(public score):0.84177。



2. 請說明你實作的 discriminative model, 其訓練方式和準確率為何? 首先將 data 106 項 feature 加入, 再加入前 6 項除了 sex 之外 5 項的平方跟立方, 把 weight 跟 bias 的 learning rate 分別設為 0.5 跟 0.1, regularization coefficient 為 0.1, 並 train10000 次, 準確率(public score):0.85737。

Machine Learning Hw2 – Report

學號: B03901026 系級:電機三 姓名:許凱傑

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。

第 2 題所提出的即為經過 feature normalization,如果相同 train 方式卻沒有經過 normalization,準確率(public score)降為 0.78243。這個結果可能是因為在 train 的時候會受到特別大的數值影響,導致 loss 的移動不會往最整理最小路 線走。

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。

在這次的作業中我實作了 regularization coefficient 為 0.1, 1.0, 10.0, 50.0, 100.0 在自己的 validation 上進行測試,發現對於結果完全沒有影響。

5. 請討論你認為哪個 attribute 對結果影響最大?

將|weight|>1的項整理如下列表格,其中上標表示是幾次項

Term	Age ¹	C_gain ¹	C_loss ¹	Age ²	C_loss ²	C_gain ³	C_loss ³
Weight	3.2498	2.2773	-1.3734	-2.9169	2.6745	1.4469	-1.1220

綜合上述我覺得影響最大的是應該是 Capital_gain 跟 Capital_loss。