Homework 2 Report - Income Prediction

學號：b04505021系級：工海三 姓名：黃廉弼

1. (1%) 請比較你實作的generative model、logistic regression的準確率，何者較佳？

|  |  |  |
| --- | --- | --- |
|  | ACC on public score | ACC on private score |
| Generative model | 0.76474 | 0.76280 |
| Logistic Regression | 0.84815 | 0.84571 |

明顯可看出Logistic Regression的Performace較Generative Model好上許多，也許是跟老師上課提及的，generative model比較適合在data量比較少的時機；Logistic Regression比較適合在data量比較多的時候，因此在這邊Logistic regression會有較佳的ACC。

1. (1%) 請說明你實作的best model，其訓練方式和準確率為何？  
   主要以上課教到的Logistic Regression實作，搭配normalization、regularization、batch normalization等技巧。  
   而Loss以cross entropy計算，Activation function則使用sigmoid function，並未使用任何optimizer(adagrad、adam…)，即僅用常數的learning rate。  
   最終ACC落在85.5%上下。
2. (1%) 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。(有關normalization請參考：<https://goo.gl/XBM3aE>)

|  |  |  |
| --- | --- | --- |
|  | ACC without normalization  (public/private) | ACC with normalization  (public/private) |
| Generative model | 0.83736/0.84250 | 0.76474/0.76280 |
| Logistic Regression | 0.79668/0.78638 | 0.84815/0.84571 |

可以看出來，在同樣的Learning Rate 以及epoch number下，Logistic Regression 中，feature normalization明顯對perfoemance 有極大的幫助。

1. (1%) 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。(有關regularization請參考：<https://goo.gl/SSWGhf> P.35)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| lambda | 0 | 1e-5 | 1e-3 | 1 | 10 |
| ACC(private) | 0.84815 | 0.83859 | 0.84559 | 0.76329 | 0.72214 |
| ACC(public) | 0.84571 | 0.84754 | 0.85196 | 0.76523 | 0.72248 |

其實不難看出，regularization對於整體performance並無太大幫助，甚至當lambda大於1後有拖累performance的狀況。

雖然說，經過shuffle，performace有所浮動在所難免，但整體趨勢看來regularization確實在此處是軍無用武之地。

1. (1%) 請討論你認為哪個attribute對結果影響最大？