

學號：B03701221 系級： 工管四 姓名：王逸庭

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

答：

logistic regression 的準確率表現較佳。

generative model: 平均準確率：0.83975

allvariable_gen_1.csv a few seconds ago by kelly wang add submission details	0.83761	0.84189
---	----------------	----------------

logistic model:平均準確率：0.852275

allvariable.csv 8 days ago by kelly wang add submission details	0.85001	0.85454
--	----------------	----------------

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

變數：全部的一次方以及部分二次方（capital-loss, hours-per-week, education-num, fnlwgt, age）

先將全部的資料標準化，再將訓練資料切成十份，取九份做訓練一份作為 validation，共做十次。

而洗牌每份訓練資料 1000 次，一次拿 50 列（份）(batch = 50)資料以 logistic regression 去訓練，來找到最好的權重。

* learning rate=0.001

allvariable_6.csv 2 days ago by kelly wang add submission details	0.85419	0.85884	<input type="checkbox"/>
--	----------------	----------------	--------------------------

準確率： $(0.85419+0.85884)/2=0.856515$

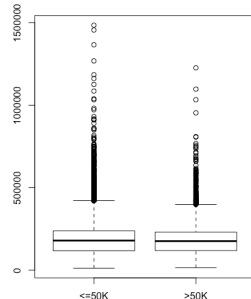
3.請實作輸入特徵標準化(feature normalization)，並討論其對於你的模型準確率的影響。

答：

allvariable_9.csv
24 minutes ago by kelly wang
no normalization

0.76231

0.76523



第二題與第三題兩個模型只有差有沒有標準化，有標準化的模型表現比為標準化的表現好很多。這是因為這次的變數域值變化差異很大，像是 fnlwgt 這個變數大部分的值都落於 20000 以上，但很多類別變數則是分佈於 0 或 1，如果未標準化容易受過大的變數影響。

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：

allvariable_8.csv
a day ago by kelly wang
lamda

0.85345

0.85835

對準確率的影響反而是變差的（差不多），這可能和訓練資料的變數次方不多有關

5.請討論你認為哪個 attribute 對結果影響最大？

如圖，因為資料以標準化過，從找到的最佳權重可以看出

-2.1077562555353704,	age	
2.2926083686568126,	capital_loss	
1.401255999421699,	fnlwgt	
-1.0081117773408363,	fnlwgt^2	
0.9303606183760685,	Federal-gov	
0.6067536708786757,	Married-spouse-absent	
-0.5475863403840436,	10th	
-0.4306879459076003,	Separated	
0.41119885274229745,	capital_gain	
-0.3874517957196766,	Prof-school	
-0.3217173018065118,	Unmarried	
0.32139224387558835,	Amer-Indian-Eskimo	
0.256450925080149,	Doctorate	
0.2369742658394748,	Farming-fishing	
0.2337731944609436,	Preschool	
-0.23274952399720727,	Priv-house-serv	
-0.21420031863643865,	Prof-specialty	

capital_loss 對於結果的影響最大，而其次 age 和 fnlwgt(包含一次和二次)影響也很大