

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

答：將助教提供的one-hot encoding全部都當作features, 然後把training data分成兩類求出各自的mean和covariance, 接著在共用covariance的情況下用條件機率搭配prior probability預測出test data的結果, 最後加上不同的scaling來比較。

Scaling	Training accuracy	Public score(accuracy)
X	0.6637	0.66499
min-max	0.7592	0.76032
z-score	0.8099	0.80983

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

答：將助教提供的one-hot encoding全部都當作features, 然後做min-max scaling, 調不同的learning rate跟epoch並使用adagrad, 接著將test data代進算出的w和b中並通過sigmoid作預測。

Learning rate	epoch	scaling	Training accuracy	Public accuracy
5e-1	2800	min-max	0.85403	0.85295
5e-1	5600	min-max	0.85338	0.85295
1e-1	2800	min-max	0.84435	0.84496
1e-1	5600	min-max	0.84810	0.84853

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：(做在discriminate model上, 並承上題的features抽取方式)有feature scaling效果會比沒scaling好, 因為在相同epoch下收斂較快；而z-score normalization效果又比min-max稍好一點, 我覺得是因為min-max會讓數值落在[0, 1]範圍內, 使得 $z=wx+b$ 時的改變量不那麼顯著, 而且標準差會變更小, 導致通過sigmoid後效果會比起z-score來得稍差。

Learning rate	epoch	scaling	Training accuracy	Public accuracy
5e-1	2800	X	0.78720	0.79742
5e-1	2800	min-max	0.85403	0.85295
1e-1	2800	min-max	0.84435	0.84496
5e-1	2800	z-score	0.85332	0.85307
1e-1	2800	z-score	0.85323	0.85319

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：(做在discriminative model上, learning rate = 5e-1, epoch = 2800, scaling = z-score normalization)若是lambda value太小, 基本上跟沒有做regularization差不多, 然後在lambda value大到一定程度時, 在public data上跟training時的accuracy也差不多。我覺得是因為我實作的model對於這些資料來說結構太簡單了, 需要更好的model來達到overfitting, 現在的model做出來的結果accuracy最多都85%而已, 可能還在underfitting, 所以做regularization會沒有什麼效。

Lambda value	Training accuracy	Public accuracy
0	0.85332	0.85307
0.01	0.85332	0.85307
100	0.85240	0.85123
10,000	0.78492	0.78391

5.請討論你認為哪個attribute對結果影響最大？

根據我自己切valid data測試的結果, 我發現capital gain對結果影響最大。在只有一次方terms的情況下我嘗試移除不同attribute, 看看各個attribute對結果的影響, 結果在移除capital gain後accuracy下降最大, 大概下降了快2%, 相對移除其他attribute最多也只下降1%左右。