

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

答：

從 training data 中隨機取出 3 萬筆 data，然後比較在 training data 中的預測準確率。但是測試下來，準確率最高還是將 training data 全取，其準確率為 0.8910

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

答：

每一次刪除一個 feature 後，再將剩下的 feature 全部拿去做 training，得到越低的預測率即表示該刪除的 feature 越重要。利用此方法得到幾個比較重要的 feature 來建構我的 model，其準確率為 0.9018

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：

有用特徵標準化的準確率：0.9018

沒用特徵標準化的準確率：0.8182

因為我有使用到某些 attribute 的 2 次式來當作 feature，所以不作特徵標準化的話，參數的設定上非常難調整，準確率也較低

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：

有用正規化的準確率：0.90181164

沒用正規化的準確率：0.90181118

兩者沒有明顯的差異

5.請討論你認為哪個 attribute 對結果影響最大？

利用 2.所提到的方法，分析出來的結果：capital_gain 為最重要的 attribute