

HW2 Decision Tree

1.Entropy

算出該dataframe的entropy，計算方式如下（其中Prob_True代表該dataframe中出現true的機率，Prob_False代表該dataframe中出現false的機率）

```
entropy = -( Prob_True * np.log2(Prob_True) + Prob_False * np.log2(Prob_False) )
```

2.Find the Best Split

Iterate through dataframe中的每個attribute，並在每個attribute中定出各種不同的threshold，再計算以各種不同的threshold split dataframe後的entropy，最後取出最小entropy的attribute和threshold作為該次split的標準。

3.Decision Tree Building

Build Decision Tree function中，每次node會利用find best split找出該次最適合split的data和threshold並將data split，再recursive build tree。

4.Classification with the MIMIC Dataset

先將24641筆testing data分割，其中index 1~16000之data作為training data，index 16001~24641之data作為validation data。利用sklearn中DecisionTreeClassifier train對training data做訓練，再將validation data套入訓練好的model，並算出其F1 score為0.43