Kiana Vang
X500: vang0951
CSCI 5521

HW4 Q3 Written Responses

A. The training and validation accuracy rates by different θ:

```
In [104]: runfile('/Users/kiavang/CSCI5521/hw4_programming/hw4.py', wdir='/Users/
kiavang/CSCI5521/hw4_programming')
Reloaded modules: MyDecisionTree
Training/validation accuracy for minimum node entropy 0.010000 is 1.000 / 0.863
Training/validation accuracy for minimum node entropy 0.050000 is 0.999 / 0.863
Training/validation accuracy for minimum node entropy 0.100000 is 0.997 / 0.865
Training/validation accuracy for minimum node entropy 0.200000 is 0.990 / 0.867
Training/validation accuracy for minimum node entropy 0.400000 is 0.979 / 0.861
Training/validation accuracy for minimum node entropy 0.800000 is 0.919 / 0.856
Training/validation accuracy for minimum node entropy 1.000000 is 0.871 / 0.840
Training/validation accuracy for minimum node entropy 2.000000 is 0.596 / 0.600
Test accuracy with minimum node entropy 0.200000 is 0.872
```

The complexity parameter θ that should be used is 0.2 with accuracy rate 0.990 and 0.867 for training and validation, respectively.

B. Given the training and validation accuracy rates, the model complexity of the Decision Tree showed that when θ is small (variance is high), the tree grew to reflect the training set accurately as shown with θ = 0.01. On the other hand, when θ is large (variance is low), the tree did not grow to reflect the training set well as shown with θ = 2.0. Therefore, the tree with the best generalization that didn't overfit or underfit the training data was the tree with the highest validation accuracy rate.

Overall, the validation accuracy rates for θ = 0.1 to 1.0 is relatively consistent around ~84-86%. The only instance where the tree underfit the training data and had poor validation accuracy was when θ = 2.0.