My strategy is prune non-leaf node form the node 1 level above each leaf node as long as the prune doesn’t hurt the performance on validation data set. For each node selection, three tests will be conducted to observe the accuracy change.

The potential order is [131072, [65535,65536],32767,16383,8191,[4095,4096],2047,1023,[511,512],[255,256],127,63……., 0]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Node ID | Test1 | | Test2 | | Test3 | |
|  | Before | Pruned | Before | Pruned | Before | Pruned |
| 131072 | 0.97 | 0.97 | 0.9 | 0.9 | 0.9 | 0.9 |
| 65535  65536 | 0.84 | 0.84 | 0.9 | 0.9 | 0.87 | 0.87 |
| 32767 | 0.87 | 0.87 | 0.97 | 0.97 | 0.9 | 0.9 |
| 16383 | 0.9 | 0.9 | 0.9 | 0.9 | 0.94 | 0.94 |
| 8191 | 0.87 | 0.87 | 0.97 | 0.84 | 0.87 | 0.87 |
| 4095  4096 | 0.94 | 0.84 | 0.94 | 0.94 | 0.9 | 0.87 |

In this case, I will prune at node is of 16383 because it will be harm for the performance. Prune at node 8191 or above will decrease the accuracy.

The out put when prune node 16383

\*\*\*\*\*\*\*\*\*\*\*\*\*Tree before pruning\*\*\*\*\*\*\*

Is PetalW == 0.2? node depth:0 node id:0

--> True:

Predict {'setosa': 24} Predicted Label: setosa

--> False:

Is PetalL == 1.5? node depth:1 node id:1

--> True:

Predict {'setosa': 8} Predicted Label: setosa

--> False:

Is PetalW == 1.3? node depth:2 node id:3

--> True:

Predict {'versicolor': 12} Predicted Label: versicolor

--> False:

Is PetalW == 0.4? node depth:3 node id:7

--> True:

Predict {'setosa': 4} Predicted Label: setosa

--> False:

Is PetalW == 0.3? node depth:4 node id:15

--> True:

Predict {'setosa': 3} Predicted Label: setosa

--> False:

Is PetalW == 0.1? node depth:5 node id:31

--> True:

Predict {'setosa': 2} Predicted Label: setosa

--> False:

Is PetalW == 1.0? node depth:6 node id:63

--> True:

Predict {'versicolor': 6} Predicted Label: versicolor

--> False:

Is PetalW == 1.4? node depth:7 node id:127

--> True:

Predict {'versicolor': 6} Predicted Label: versicolor

--> False:

Is PetalW == 1.5? node depth:8 node id:255

--> True:

Is PetalL == 5.1? node depth:9 node id:512

--> True:

Predict {'virginica': 1} Predicted Label: virginica

--> False:

Is PetalL == 5.0? node depth:10 node id:1025

--> True:

Predict {'virginica': 1} Predicted Label: virginica

--> False:

Predict {'versicolor': 8} Predicted Label: versicolor

--> False:

Is PetalW == 1.2? node depth:9 node id:511

--> True:

Predict {'versicolor': 5} Predicted Label: versicolor

--> False:

Is PetalW == 1.1? node depth:10 node id:1023

--> True:

Predict {'versicolor': 2} Predicted Label: versicolor

--> False:

Is PetalW == 1.6? node depth:11 node id:2047

--> True:

Is PetalL == 5.8? node depth:12 node id:4096

--> True:

Predict {'virginica': 1} Predicted Label: virginica

--> False:

Predict {'versicolor': 2} Predicted Label: versicolor

--> False:

Is PetalW == petal\_width? node depth:12 node id:4095

--> True:

Predict {'species': 1} Predicted Label: species

--> False:

Is PetalW == 0.6? node depth:13 node id:8191

--> True:

Predict {'setosa': 1} Predicted Label: setosa

--> False:

Is PetalW == 0.5? node depth:14 node id:16383

--> True:

Predict {'setosa': 1} Predicted Label: setosa

--> False:

Predict {'virginica': 32} Predicted Label: virginica

Accuracy on test = 0.9

\*\*\*\*\*\*\*\*\*\*\*\*\*Tree after pruning\*\*\*\*\*\*\*

Is PetalW == 0.2? node depth:0 node id:0

--> True:

Predict {'setosa': 24} Predicted Label: setosa

--> False:

Is PetalL == 1.5? node depth:1 node id:1

--> True:

Predict {'setosa': 8} Predicted Label: setosa

--> False:

Is PetalW == 1.3? node depth:2 node id:3

--> True:

Predict {'versicolor': 12} Predicted Label: versicolor

--> False:

Is PetalW == 0.4? node depth:3 node id:7

--> True:

Predict {'setosa': 4} Predicted Label: setosa

--> False:

Is PetalW == 0.3? node depth:4 node id:15

--> True:

Predict {'setosa': 3} Predicted Label: setosa

--> False:

Is PetalW == 0.1? node depth:5 node id:31

--> True:

Predict {'setosa': 2} Predicted Label: setosa

--> False:

Is PetalW == 1.0? node depth:6 node id:63

--> True:

Predict {'versicolor': 6} Predicted Label: versicolor

--> False:

Is PetalW == 1.4? node depth:7 node id:127

--> True:

Predict {'versicolor': 6} Predicted Label: versicolor

--> False:

Is PetalW == 1.5? node depth:8 node id:255

--> True:

Is PetalL == 5.1? node depth:9 node id:512

--> True:

Predict {'virginica': 1} Predicted Label: virginica

--> False:

Is PetalL == 5.0? node depth:10 node id:1025

--> True:

Predict {'virginica': 1} Predicted Label: virginica

--> False:

Predict {'versicolor': 8} Predicted Label: versicolor

--> False:

Is PetalW == 1.2? node depth:9 node id:511

--> True:

Predict {'versicolor': 5} Predicted Label: versicolor

--> False:

Is PetalW == 1.1? node depth:10 node id:1023

--> True:

Predict {'versicolor': 2} Predicted Label: versicolor

--> False:

Is PetalW == 1.6? node depth:11 node id:2047

--> True:

Is PetalL == 5.8? node depth:12 node id:4096

--> True:

Predict {'virginica': 1} Predicted Label: virginica

--> False:

Predict {'versicolor': 2} Predicted Label: versicolor

--> False:

Is PetalW == petal\_width? node depth:12 node id:4095

--> True:

Predict {'species': 1} Predicted Label: species

--> False:

Is PetalW == 0.6? node depth:13 node id:8191

--> True:

Predict {'setosa': 1} Predicted Label: setosa

--> False:

Predict {'virginica': 32, 'setosa': 1} Predicted Label: virginica

Accuracy on test = 0.9