

TDT4171 - Methods in AI

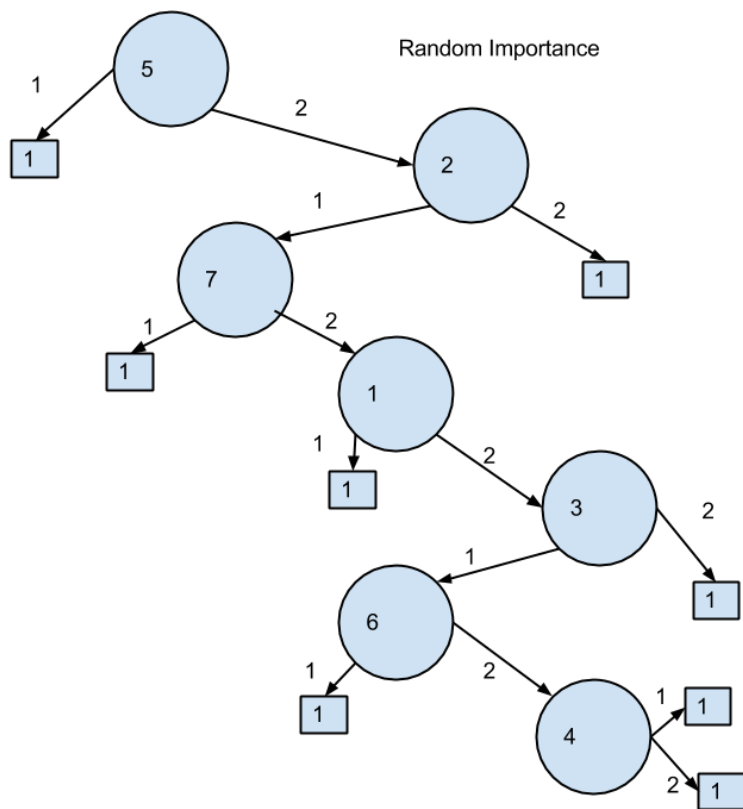
Exercise 4

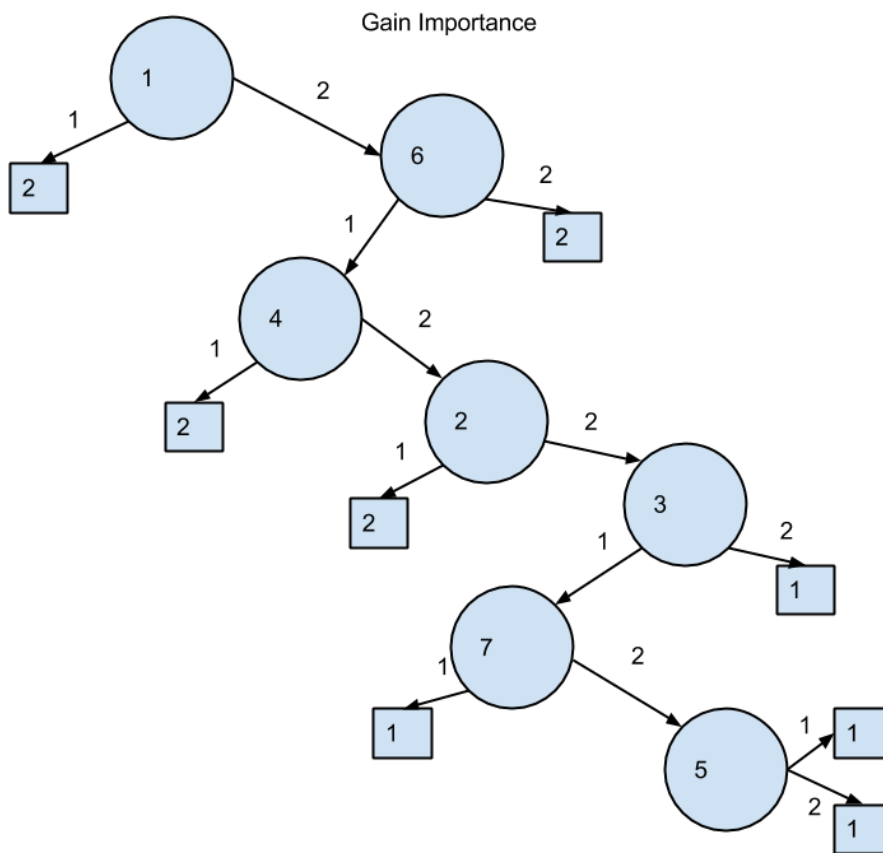
Magnus Kirø

- *What can you conclude about the results you have obtained? Which Importance is better, and why?*
 - The gain importance function provides an overall better accuracy for the decision tree. At times the results are not that good but over time it will provide better results than with random importance. We can see that the random importance will get better results at times, but that depends on the random choice of attribute key. The random importance function chooses a random attribute as the root node. The gain importance function chooses the attribute with the highest gain. Depending on the attribute set the gain value will vary.
- *Do you get the same result if you run the random Importance several times?*
 - The random importance function gives different answers for each run. This is to be expected as the split attribute is chosen randomly.
- *What happens when you run the learner based on Information Gain several times?*
 - When I run the Information gain several times I get different trees. This is not to be expected, but if the calculated gain is equal for two nodes, the choice of a root node happens randomly. This might change the tree from time to time. By running information gain several times we should see that generally this provides better results than random importance.

Tree Figures:

Number in node is the attribute.





Example Terminal output:

Decision tree, importance implementation exercise

Training tree:

```

-5
-2
-7
-1
-3
-6
-4
L:1
-4
L:1
-6
  
```

L:1
-3
L:1
-1
L:1
-7
L:1
-2
L:1
-5
L:1

Results:
correct / total
21 / 28
percentage: 0.75

Training tree:

-1
-6
-4
-2
-3
-7
-5
L:1
-5
L:1
-7
L:1
-3
L:1
-2
L:2
-4
L:2
-6
L:2
-1
L:2

Results:

correct / total

10 / 28

percentage: 0.357142857143

Finished