

# TDT4171 - Assignment 4

Aleksander Hansen

April 2016

## Discussion of findings

Based on the results it there doesn't seem to matter much which importance function to choose performance wise. The random function seems to be fluctuating around a mean that is equal to the gain functions performance. I've gotten anything from 18/28 to 23/28 correct from the random function, while the gain function stays constant at 20/28 correct when running the program several times. If stability is an issue then the gain function is better.

Now, these are just my results, and I suspect the gain function is implemented wrong as one would expect it to perform better than the random one. I haven't been able to figure out if and why it's wrong. It could also just be that the training data is not that good.

## Results

Output from console:

Random importance

Correct: 22/28

Tree structure:

```
value: 5
  value: 0
    value: 1
    value: 3
      value: 1
        value: 2
          value: 1
          value: 2
        value: 6
          value: 4
            value: 1
            value: 2
          value: 1
      value: 1
    value: 1
```

value: 1

Information gain importance

Correct: 20/28

Tree structure:

value: 0

value: 1

value: 4

value: 2

value: 1

value: 1

value: 2

value: 6

value: 3

value: 5

value: 1

value: 2

value: 1

value: 1

value: 2