

TDT4171 Assignment 4

Fredrik Veidahl Aagaard

March 2019

1 Implementation

The IMPORTANCE and DECISION-TREE-LEARNING algorithms were implemented using theory from Chapter 18 using Python3.7 and a few Numpy functions.

2 Tree structures

The tree structures I got using random importance and information gain importance are given in Listing 1 and Listing 2 respectively.

3 Results

I classified all examples in the test set to calculate the accuracy of the two importance methods. I did this 100 times for both trees, and got an average of 23.56/28 correct using random importance and 26/28 correct using information gain importance.

4 Discussion

We see that the information gain importance gave the best average results. However, the random importance did sometimes get more than 26/28 correct. Out of 100 I got 28/28 correct 6 times.

The results makes sense as the random tree will change each time we run the algorithm, which makes the accuracy random, and we will seldom get more than 26/28 correct. The information gain tree will be the same each time we run the algorithm, and in this case we will always get 26/28 correct.

```

TREE STRUCTURE USING RANDOM IMPORTANCE
Attribute 3
Attribute 7
  Attribute 5
    Attribute 1
      Class 1
      Attribute 2
        Class 1
        Class 2
    Attribute 1
      Class 1
      Attribute 6
        Attribute 2
          Class 1
          Class 2
        Attribute 4
          Attribute 2
            Class 1
            Class 2
          Class 2
    Attribute 2
      Class 1
      Attribute 5
        Attribute 4
          Attribute 6
            Class 1
            Class 2
          Attribute 1
            Class 1
            Class 2
        Attribute 1
          Class 1
          Class 2
    Attribute 5
      Attribute 2
        Attribute 1
          Class 1
          Class 2
        Class 1
      Attribute 1
        Class 1
        Attribute 6
          Attribute 4
            Attribute 2
              Class 2
              Class 1
            Attribute 7
              Attribute 2
                Class 2
                Class 1
              Attribute 2
                Class 2
                Class 1
          Attribute 2
            Class 2
            Class 1
    Attribute 2
      Class 2
      Class 1

```

Listing 1: Tree structure using random importance

```

TREE STRUCTURE USING INFORMATION GAIN IMPORTANCE
Attribute 1
Class 1
Attribute 5
  Attribute 3
    Attribute 2
      Class 1
      Class 2
    Attribute 2
      Class 2
      Class 1
  Attribute 6
    Attribute 2
      Attribute 3
        Class 1
        Class 2
      Attribute 3
        Class 2
        Class 1
    Attribute 4
      Attribute 2
        Attribute 3
          Class 1
          Class 2
        Attribute 3
          Class 2
          Class 1
      Attribute 2
        Class 2
        Attribute 3
          Class 2
          Class 1

```

Listing 2: Tree structure using information gain importance