Introduction to Intelligent Systems - Lab 4

Tom Apol (s2701650)
Nielis Brouwer (s3850706)
Group: 23

October 8, 2018

Contents

1 Assignment 1

The smallest height of our binary decision tree is 3 (with the root being height 0). This is because our number of desired leaves is 5, the maximum number of leaves on height h is 2^h, and $2^2 < 5 \leq 2^3$. Our first decision query would either be 'dorsal fin' or 'fluke visible when diving', since these are the only two queries for which we have data from every whale category.

The following series of query choices are based on whether that query divides the remaining whale categories and whether it has data for every remaining category. (This, for example, excludes the 'blows water' and 'tusk' queries entirely.) To keep the tree as short as possible, we also chose each query such that they split each group in approximately half. The resulting decision tree is just one of multiple possible outcomes of this algorithm.

Below is a truth-table summarising the responses of each whale category with respect to the queries for the decision tree, where `1` denotes 'True', `0` denotes 'False' and `?` denotes 'Not defined'.

Whale type:	dorsal fin	*fluke visible	large size whale	big fluke	blows water	tusk
Killer whale	1	0	0	0	1	?
Beluga whale	0	0	?	?	?	?
Narwhal whale	0	1	0	?	?	1
Bowhead whale	0	1	1	?	?	?
Blue whale	1	1	1	1	?	?

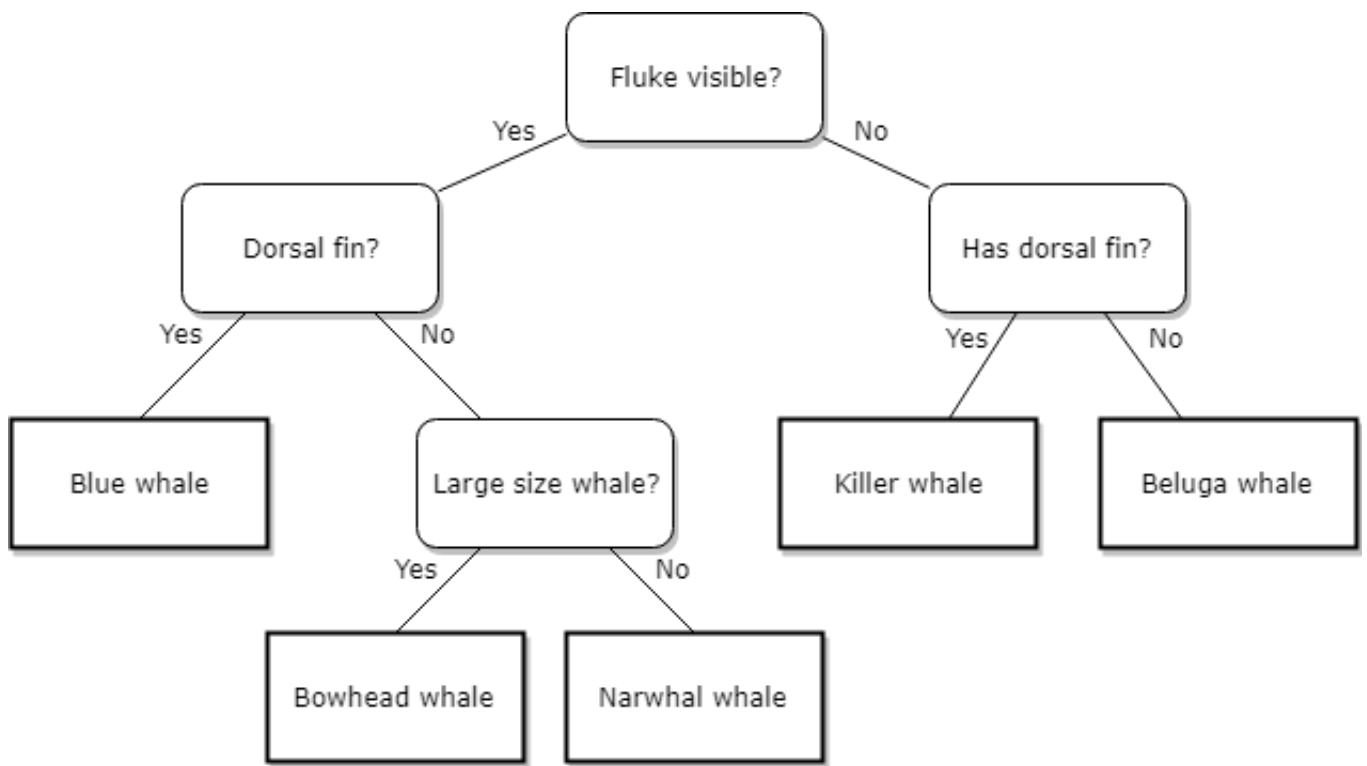* fluke visibility when diving

Figure 1: Decision tree whale types

2   Assignment 2

Since the AEX index is the weighted average of the Dutch stock market, it can be seen as the prototype of the entire Dutch stock market. In this case, it is the index with the lowest cluster distance between any other index. Due to the form of the dendrogram, we can surmise that for any index, the smallest euclidean distance between themselves and any other index or sub-cluster, is that between themselves and the largest sub-cluster they are not yet part of. This is rather intriguing, as we would expect it to be more likely for multiple smaller clusters of indices to make up the dendrogram.
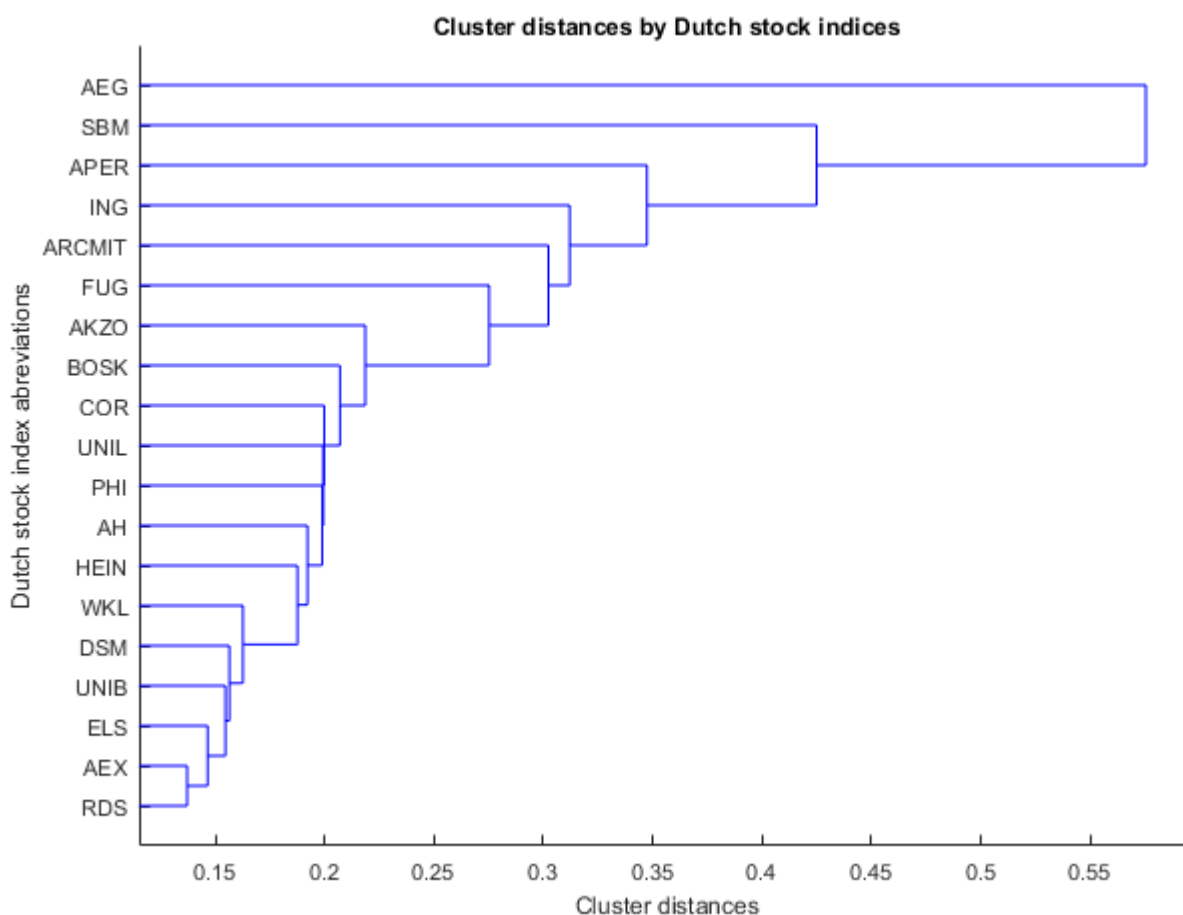

Figure 2: Cluster distances of Dutch stock indexes. For readability's sake, we've rotated the plot by 90 degrees.

Since the AEX is the weighted average of all other indices, we were curious to what degree this influences the dendrogram. Therefore we included the following dendrogram that excludes this index. Apart from some change in sub-clusters and specific distances, the general shape, and thus the aforementioned property, remained the same.
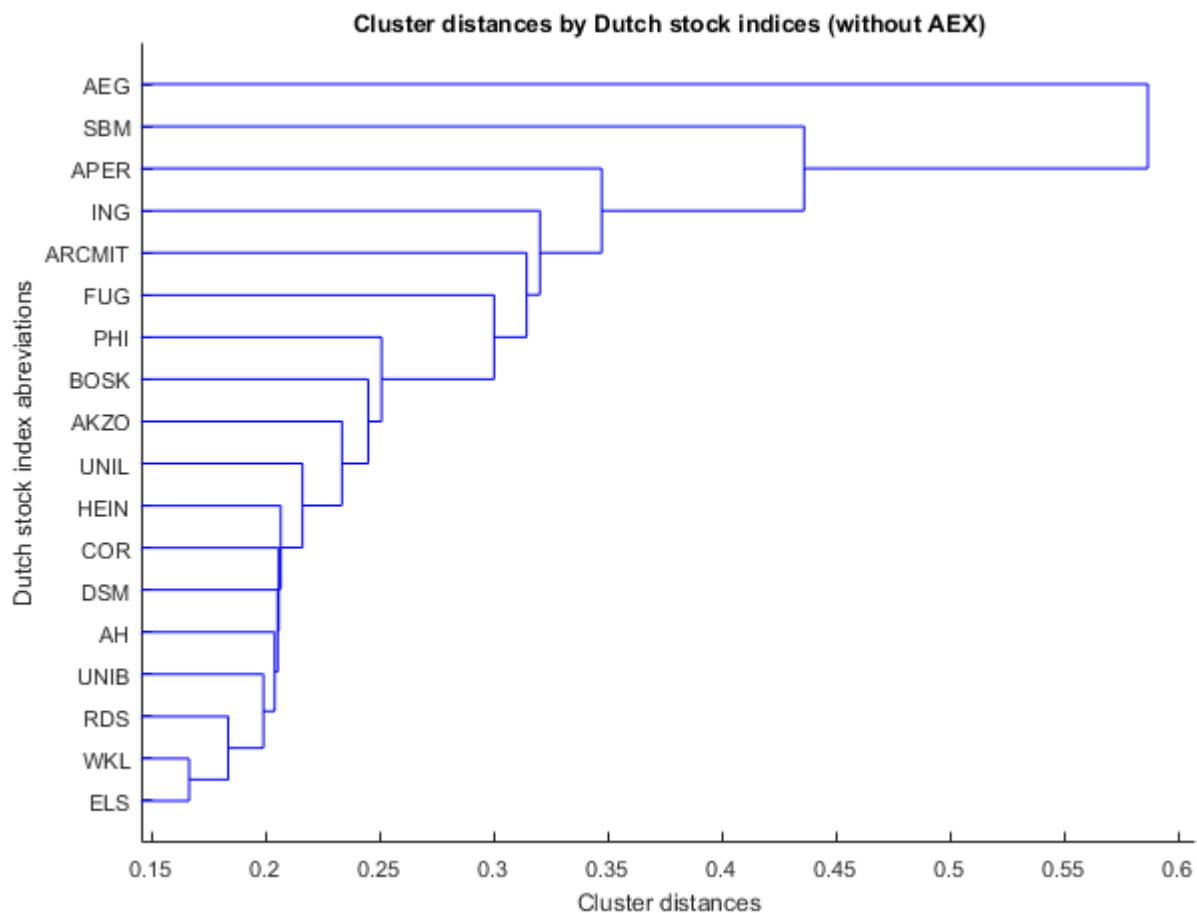

Figure 3: Cluster distances of Dutch stock indexes, excluding the AEX. For readability's sake, we've rotated the plot by 90 degrees.

See the Appendix for the corresponding code.

3   Appendix

```
 1  close all;
 2  load('lab_4_data/dataAEX.mat');
 3  load('lab_4_data/labelsAEX.mat');
 4
 5  aex_linkage = linkage(data);
 6  dendrogram(aex_linkage,'Labels', labels,'Orientation','right');
 7  ylabel('Dutch_stock_index_abreviations');
 8  xlabel('Cluster_distances');
 9  title('Cluster_distances_by_Dutch_stock_indices');
10
11  non_aex_linkage = linkage(data(1:18, :))
12  figure;
13  dendrogram(non_aex_linkage, 'Labels', labels(1:18), ...
14      'Orientation', 'right');
15  ylabel('Dutch_stock_index_abreviations');
16  xlabel('Cluster_distances');
17  title('Cluster_distances_by_Dutch_stock_indices_(without_AEX)');
```