

# Introduction to Intelligent Systems: Assignment 4

## Bayes classification and Normal Distribution

Damiano Melotti (S3838706)  
Amadeus Beckmann (S3839508)  
Group 55

October 7, 2018

### 1 Assignment 1

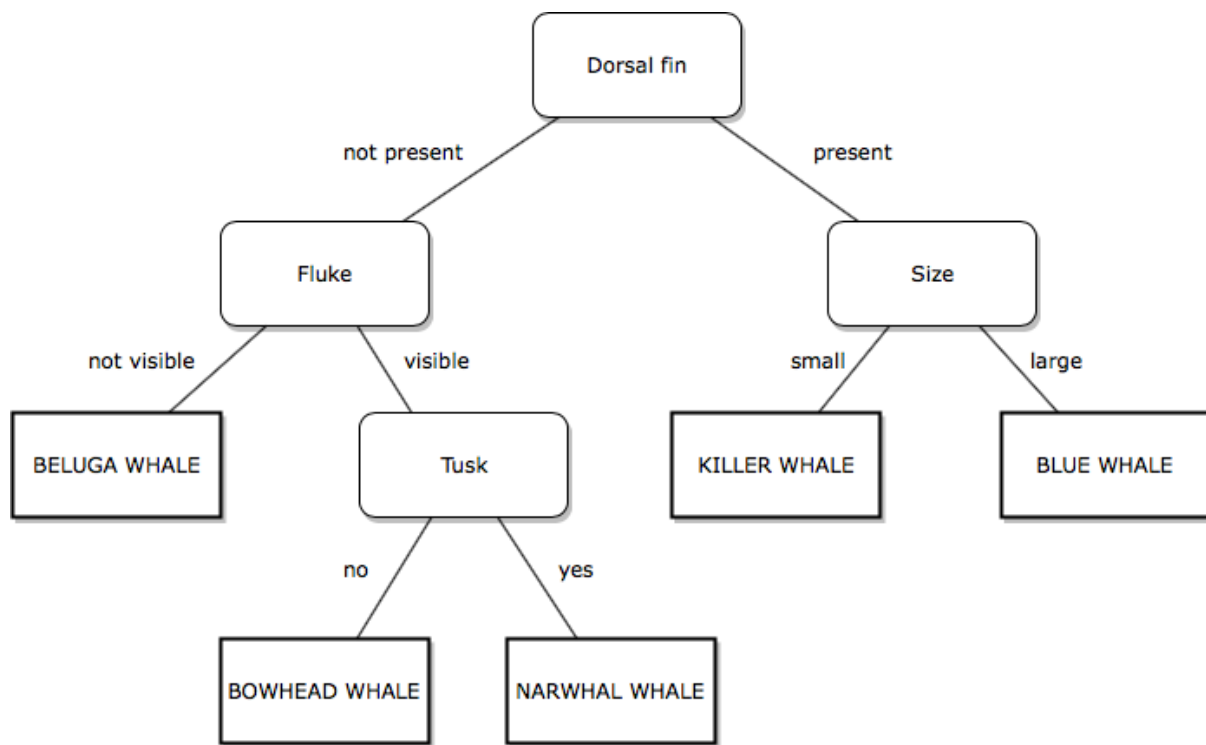


Figure 1: Binary decision tree for whale spotters

Figure 1 shows a binary decision tree that can be used to distinguish **killer whales**, **beluga whales**, **narwhals**, **bowhead whales**, and **blue whales**. Using easily identifiable characteristics, it can be determined with minimal effort which specimen is in question.

Our feature vector was: <size, dorsal fin, visible fluke, tusk>.

Size: [small; medium; large]

Dorsal fin: [present, not-present]

Visible fluke: [visible; not-visible]

Tusk: [yes; no]

Considering the characteristics the marine biologist provided, we came up with the following schematization:

**Killer whale:**

1. Fluke not visible when diving;
2. Small size (6-8m);
3. Tall and pointed dorsal fin is often clearly visible;
4. Blows water quite often.

Vector: <s,pr,nv,?>

**Beluga whale:**

1. No fluke when diving;
2. No dorsal fin.

Vector: <?,np,nv,?>

**Narwhal:**

1. Very small;
2. Long tusk;
3. Fluke is clearly visible;
4. No dorsal fin.

Vector: <s,np,v,y>

**Bowhead whale:**

1. Large size of 20m;
2. Fluke is visible;
3. No dorsal fin.

Vector: <l,np,v,?>

**Blue whale:**

1. Very large size of 30m;
2. Impressive fluke;
3. Small dorsal fin;

Vector: <l,p,v,?>

Our first thought was to split the whales into two equal groups. We had to take into account that there are unknown features in some of the animals, so only two characteristics can be considered, which apply to half of the whales: the possession of a dorsal fin and the visibility of the fluke. However, we've made the possession of a **dorsal fin** the first decision.

The only two whales that have a dorsal fin are the killer whale and the blue whale. They are pretty easy to differentiate: one is 6-8m long, the other 30m. So the choice of the decision criterion for these two mammals is their **size**.

The remaining three whales, which do not have a dorsal fin are also pretty easy to distinguish, because one of them has no visible **fluke**: that is the beluga whale, which means only the narwhal and the bowhead whale remain.

Finally, it is not difficult to recognize the remaining two whales: the narwhal has a long **tusk** and the bowhead whale does not.

With this binary decision tree not only the height of the tree is minimal, but also the characteristics to determine are well recognizable.

## 2 Assignment 2

### 2.1 Dendrogram

We've created the dendrogram following the given indications. To calculate the dissimilarity between the elements of our data set and generate the matrix we used the Euclidean distance. Then we created a hierarchical binary cluster tree with the default linking function offered in MATLAB, and we generated the dendrogram.

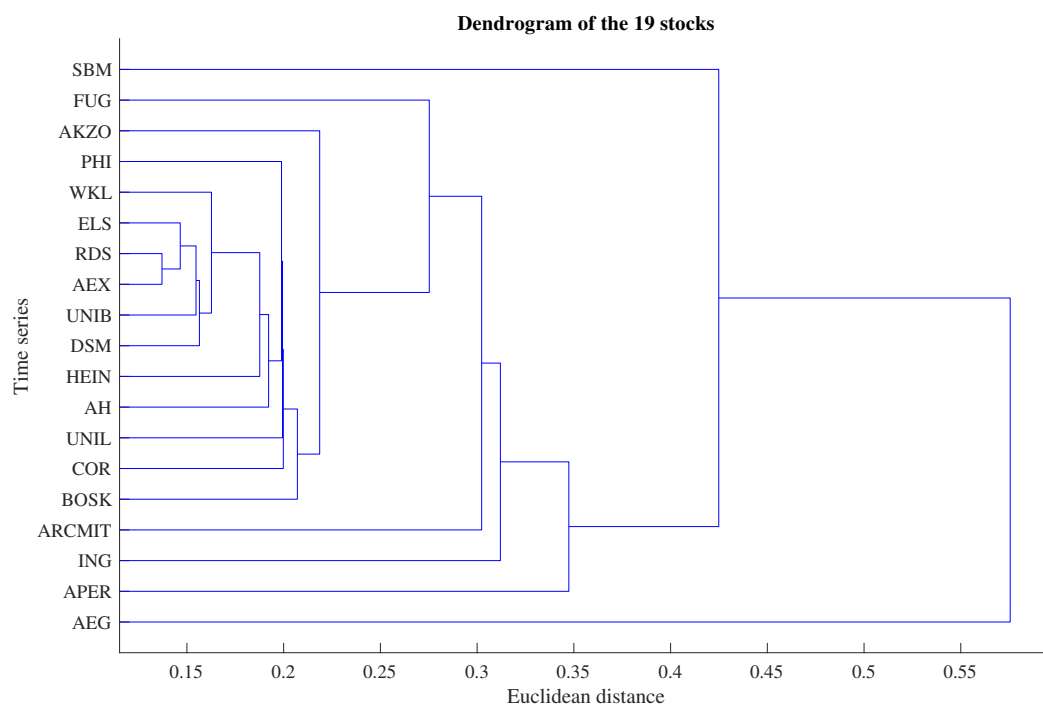


Figure 2: Dendrogram of 19 stocks

### 2.2 Considerations

When looking at the dendrogram, it is noticeable that the largest Euclidean distance is almost 600% greater than the smallest. We suspected that the dendrogram visualizes the similarity between the market values: a small Euclidean distance means a large correspondence between the changes of the market values and vice versa. To confirm our thesis, we have selected two examples, one with a small and one with a large Euclidean distance, and plotted their market values.

Of course, the two graphs in figure 2.2 are not equal. They still have a lot of dissimilarities. However, if you compare it with figure 2.2, you can see that we were right with our assumption. While AEX in figure 2.2 has comparatively low amplitudes, the AEG graph behaves more remarkably. Almost all values are far below or above the corresponding AEX market value. The spike at  $\sim 110$  also intensifies this.

Furthermore, while the market values in figure 2.2 lie between  $-0.04$  and  $0.04$ , they are rather between  $-0.075$  and  $0.1$  in figure 2.2.

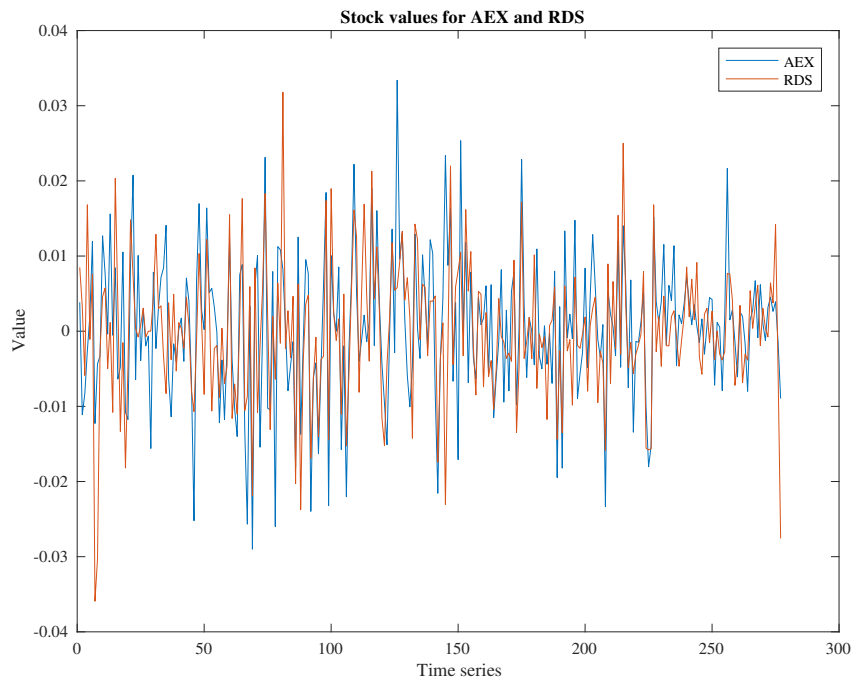


Figure 3: AEX and RDS

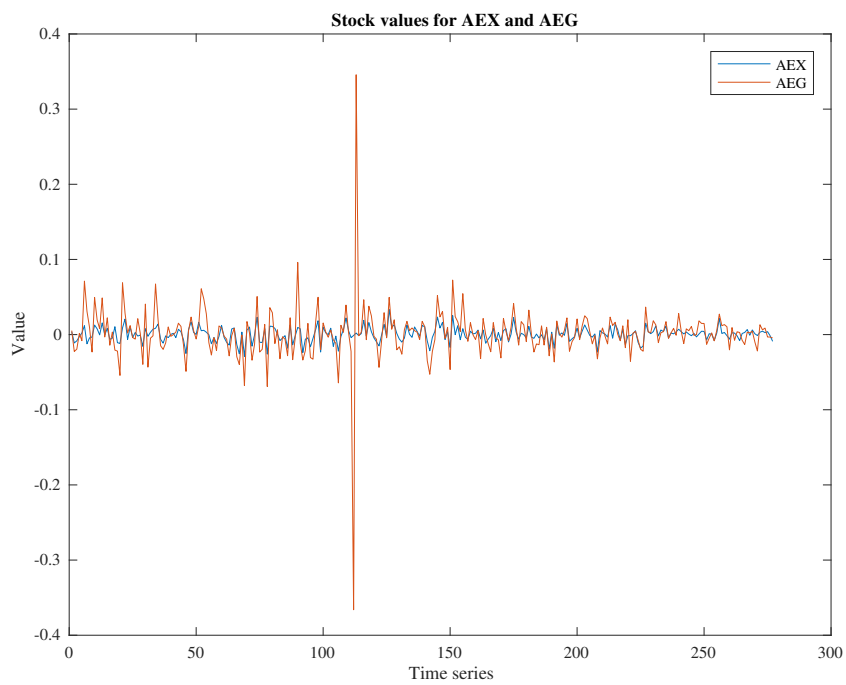


Figure 4: AEX and AEG

## 2.3 Code (Dendrogram)

```
load('lab_4_data/dataAEX.mat');
load('lab_4_data/labelsAEX.mat');

% euclidean (default) distances between the time series
distances = pdist(data);
% to visualize better the matrix
distance_matrix = squareform(distances);

% to create a hierarchical binary cluster tree
tree = linkage(distances);
% to compute the best order of leaves
leafOrder = optimallinkage(tree, distances);

% we use introduce the labels in the plot
[H,T] = dendrogram(tree, 'Reorder', leafOrder, 'Labels', labels, 'Orientation',
'right');
title('Dendrogram of the 19 stocks');
xlabel('Euclidean distance');
ylabel('Time series');
set(gca, 'fontsize', 9.5, 'fontname', 'Times New Roman');
```

## 2.4 Code (Graphs)

```
load('lab_4_data/dataAEX.mat');
load('lab_4_data/labelsAEX.mat');

indexAEX = find(strcmp(labels, 'AEX'), 1);
indexRDS = find(strcmp(labels, 'RDS'), 1);
indexAEG = find(strcmp(labels, 'AEG'), 1);

plot(data(indexAEX, :));
hold on;
plot(data(indexRDS, :));
legend('AEX', 'RDS');
title('Stock values for AEX and RDS');
xlabel('Time series');
ylabel('Value');
set(gca, 'fontsize', 9.5, 'fontname', 'Times New Roman');

figure(2);
plot(data(indexAEX, :));
hold on;
plot(data(indexAEG, :));
legend('AEX', 'AEG');
title('Stock values for AEX and AEG');
xlabel('Time series');
ylabel('Value');
set(gca, 'fontsize', 9.5, 'fontname', 'Times New Roman');
```