

# TI2736-C Assignment 4

Joost Pluim, jwpluim, 4162269  
Pascal Remeijns, premeijns, 4286243

March 17, 2016

## **Hierarchical clustering**

### **1 Question 4**

#### **1.1 Question 4.1**

At the leaves we find every single node. If we set  $k$  to 1 we find at the root of the tree the cluster with all nodes.

#### **1.2 Question 4.2**

$k$  sets the number of clusters, or the number of colours the tree has.

### **2 Question 5**

#### **2.1 Question 5.1**

We see that not every line is one cluster. Because we use the mean distance, at some point the middle and lower line are merged together.

### **3 Question 7**

#### **3.1 Question 7.1**

You clearly see that all points on a line should belong together in one of the three clusters, and they also do. Now the distance between two points on a line is always smaller than the distance between two points in different clusters on different lines.

### **4 Question 8**

#### **4.1 Question 8.1**

Now the result isn't better because the two furthest outliers negatively influence the result now, because those are the points with the longest minimum distance to whatever other point/cluster. This means they will both have their own cluster. This means they will always be handled in the end of the algorithm, and therefore points are never added to their cluster.

***$k$ -means clustering***

### **5 Question 1**

#### **5.1 Question 1.1**

One way is to say that you search for the  $k$  points which are furthest away from each other. Disadvantage is that finding these points is quite high complexity, advantage is that approach the clusters from the outside, so the initial clusters will already be quite optimal.

## **6 Question 5**

### **6.1 Question 5.1**

A good choice for  $k$  would therefore be  $k = 10$ .

### **6.2 Question 5.2**

If we set  $k$  to the number of points, this means every point will have its own cluster. This means the RSS will be 0 for every cluster, so the sum is also 0.

## **References**