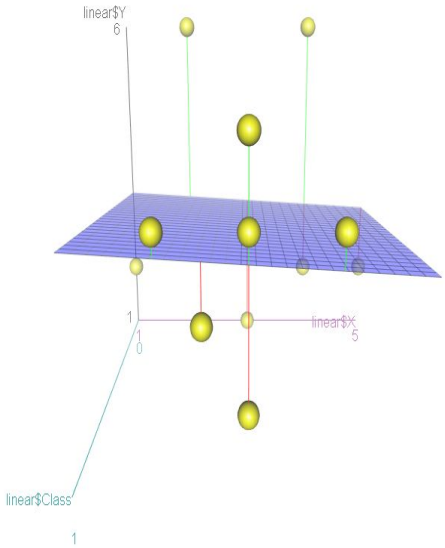
HOMEWORK 11

KENIGBOLO MEYA STEPHEN

1. Load the dataset from here (or csv version). The data points belong to two classes - positives and negatives. Notice that two classes are not linearly separated in the original 2D feature space. However, by applying the "kernel trick" we can map the original feature space into a high dimensional one where the classes would be linearly separable. Try to come up with new feature(s) based on X and Y, such that the given points would be separable. (Hint)

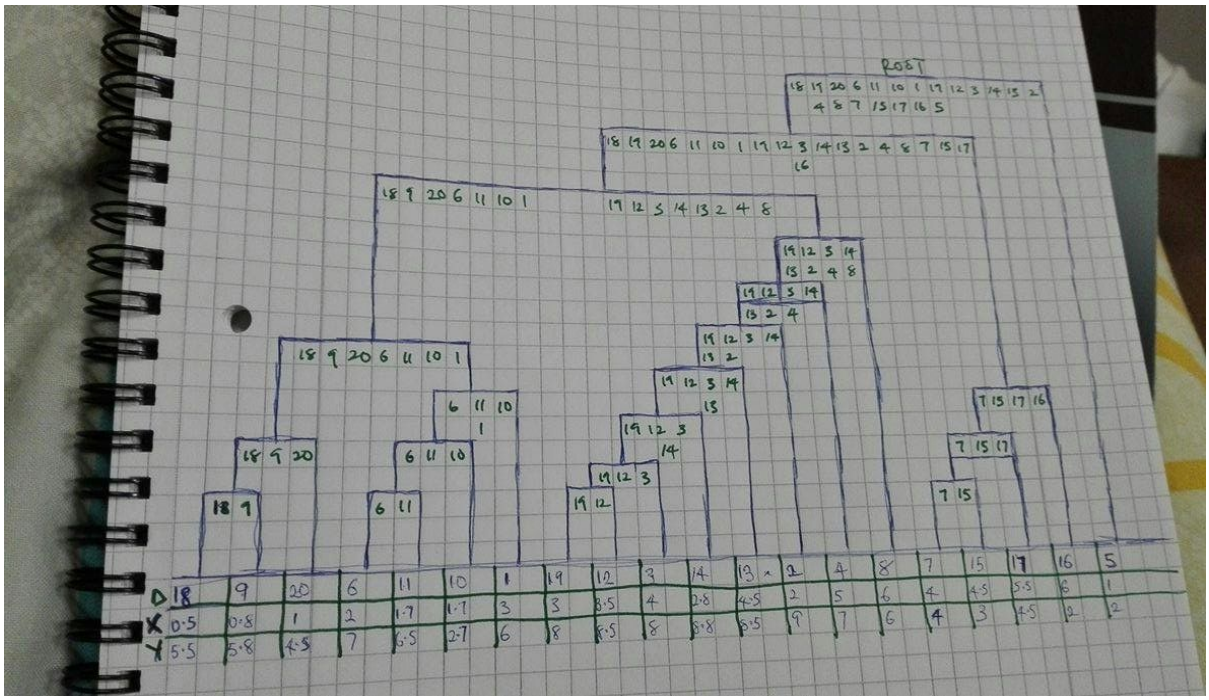


For this task I used the scatterplot3D package from library(Rcmdr) in order to split the cluster.

2. Load dataset from this Excel file (or csv version). Your task is to simulate hierarchical clustering:

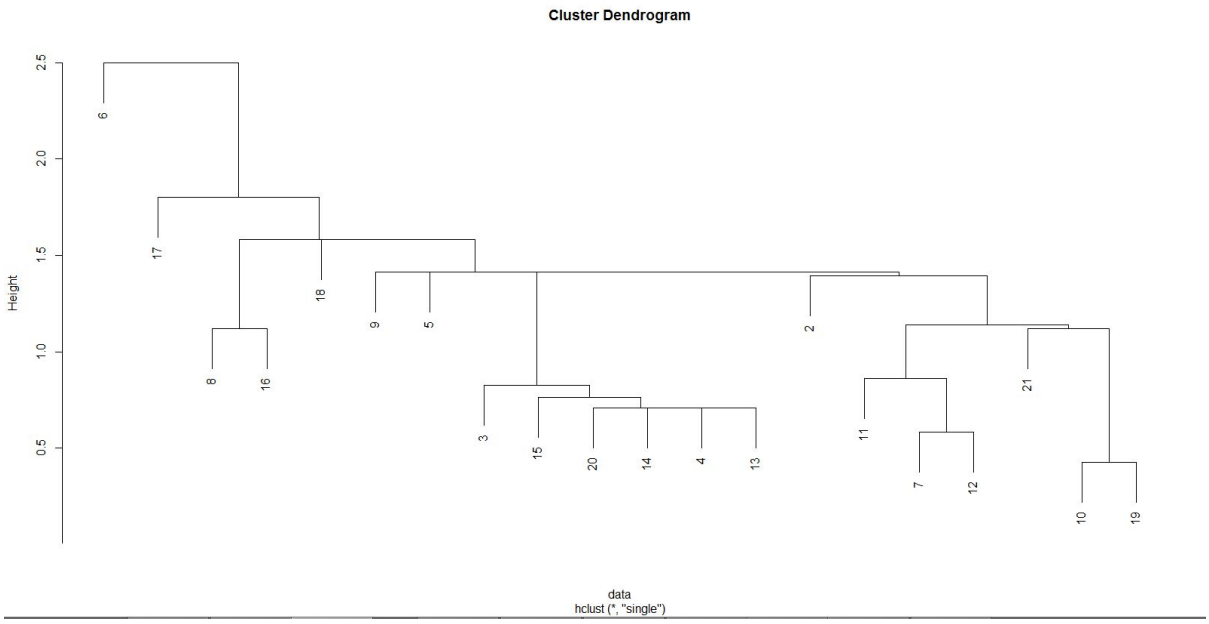
1. Single link (min distance) clustering
2. Complete link (max distance) clustering

Use common sense, no need to calculate ALL distances. Draw by hand to save time…

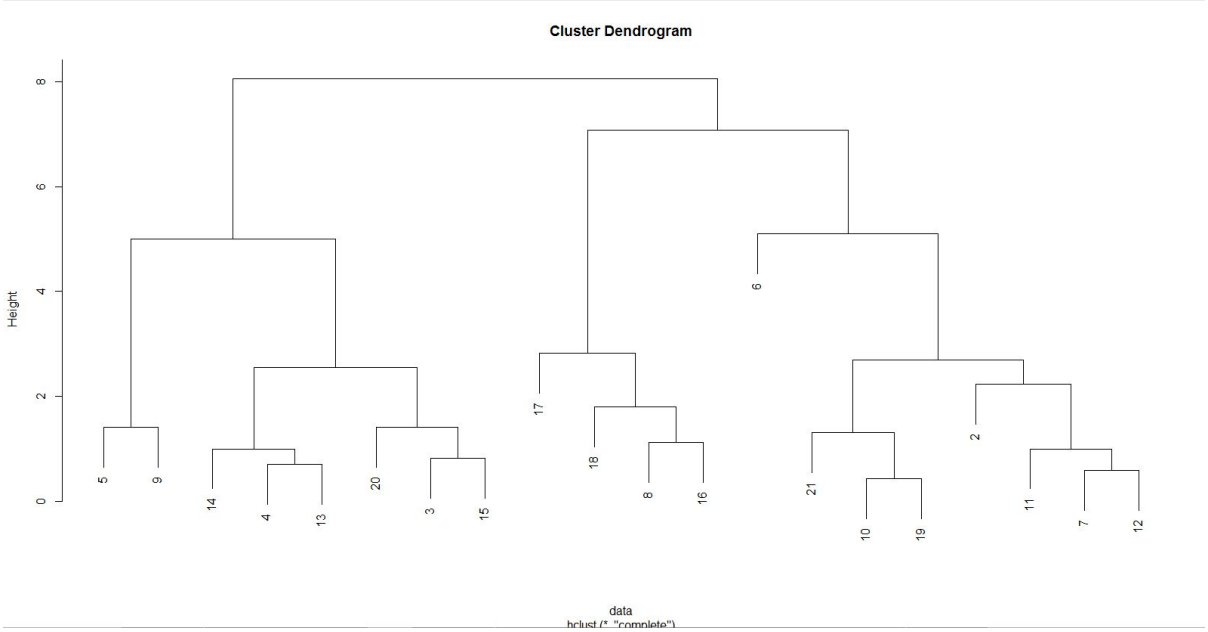


N/B -> All simulation pictures can be found in the folder “Simulation” as proof that I do understand how the algorithm works. I resorted to not putting it here in order to ensure that the pdf isn’t just folded with pictures.

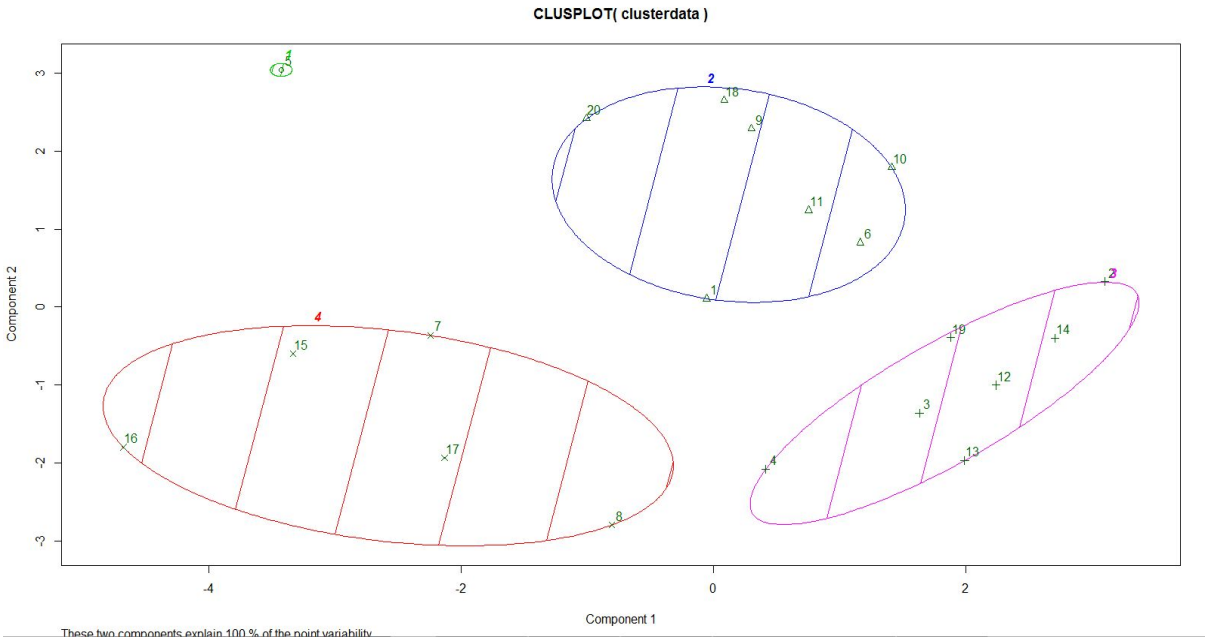
Single link



Complete Link



3. Use the same data, and use first 4 points as K cluster centers for K-means. Simulate the K-means (using Euclidean distance). Again, use common sense and approximate distances where needed. When in serious doubt, you can rely on more precise calculations.



First four points as K cluster centers

I used R for this task. The file can be found attached in the zip file