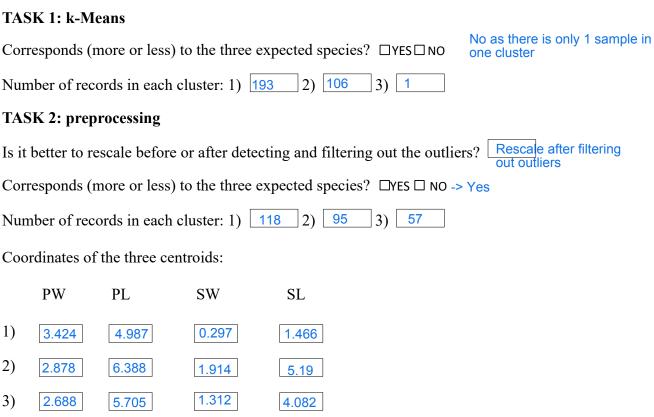
Submission sheet Assignment III ASK 1: k-Means



TASK 3: choice of k

Which K corresponds to the best clustering? (using the Davies-Boulding index).

k=2 -> score: 0.353

TASK 4: Hierarchical clustering

Using SingleLink, how many records are included in each of the two top clusters?

Cluster 1: 175

Cluster 2: 95

Which approaches produce a (more or less) correct clustering corresponding to the three species, if any?

SingleLink: No -> since individual points get added to the big clusters at later stages

CompleteLink: Yes, there are 3 distinct clusters which are merged in the last 2 steps

AverageLink: Yes, there are 3 distinct clusters which are merged in the last 2 steps

TASK 5: DB-Scan

How many clusters does DB-SCAN find with eps=1, min_samples=5? 2
Can you give a value for epsilon leading to two clusters (plus noise)?
K-DISTANCES:
Which K did you use? 5
According to the k-distances plot, what value(s) of epsilon would you consider as a parameter to DB-Scan and why?
I would consider e=0.25 as a parameter to DB scan, since after wards there is a steep incline i the plot