**Submission sheet**

Assignment III

# TASK 1: k-Means

Corresponds (more or less) to the three expected species? YES NO

Number of records in each cluster: 1) 103 2) 112 3) 85

# TASK 2: preprocessing

Is it better to rescale before or after detecting and filtering out the outliers? After.

Corresponds (more or less) to the three expected species? YES NO

Number of records in each cluster: 1) 103 2) 100 3) 89

Coordinates of the three centroids:

PW PL SW SL

1) [0.5897358 0.41279922 0.69850187 0.70067416 0.79044079 0. ] 2)[0.22881134 0.64288729 0.0907767 0.0815534 0.52157874 0. ]. 3) [0.54648649 0.38391304 0.6245 0.606 0.29531986 0.]

# TASK 3: choice of k

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

# TASK 4: Hierarchical clustering

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

Cluster 1:

Cluster 2:

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

SingleLink:

CompleteLink:

AverageLink:

# TASK 5: DB-Scan

How many clusters does DB-SCAN find with eps=1, min\_samples=5?

Can you give a value for epsilon leading to two clusters (plus noise)?

K-DISTANCES:

Which K did you use?

According to the k-distances plot, what value(s) of epsilon would you consider as a parameter to DB-Scan and why?