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) ☐ 2) ☐ 3) ☐

TASK 2: preprocessing

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

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

Number of records in each cluster: 1) ☐ 2) ☐ 3) ☐

Coordinates of the three centroids:

	PW	PL	SW	SL
1)	☐	☐	☐	☐
2)	☐	☐	☐	☐
3)	☐	☐	☐	☐

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?