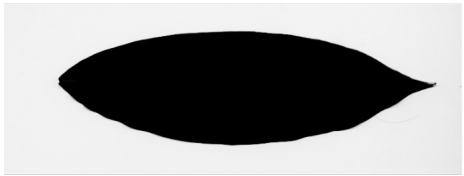
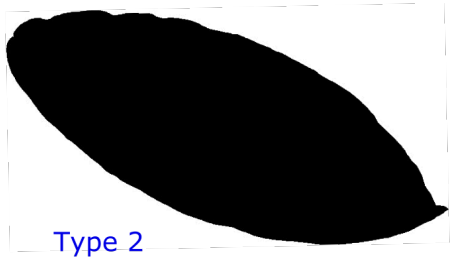
Problem



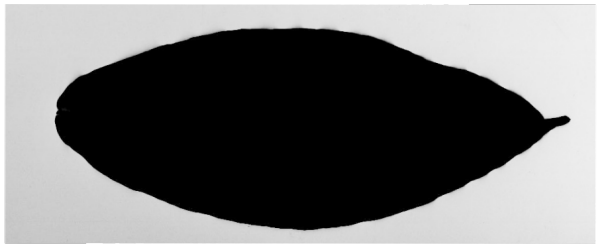
Type 1

Type 3



Type 2

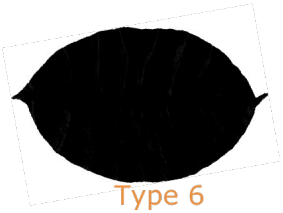
Type 6



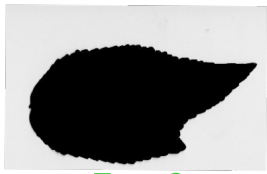
Type 2

Type 1

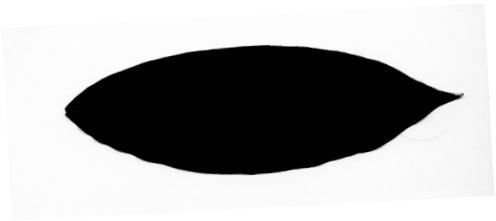
Type 4



Type 6



Type 3



Species

We need features

Variable 1

- For different leaves the length of curvature will be different
- So, we wish to take a vector that will be of same length for all leaves.
- So we break the range of the curvature i.e. (0,1) into 20 parts and take the number of curvatures lying in the 20 intervals.

Variable 2

- As we can see, variable 1 alone is not enough to classify the leaves.
- Hence we introduce 2 new variables.

Variable 3

- The number of Inward and outward protrusions, or peak, often vary for different leaves.
- Hence we consider that feature for our analysis.

Dendrogram

- A dendrogram is a is a tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering.
- It first take the 2 points at shortest distance, and take their average. It then discards the two points and take their average as the new cluster.

Variable 4

- Another classifier that we thought of taking into account is the thickness or thinness of the leaf.
- For this, we will use the idea of Principal Component Analysis.
- We will the consider the two eigen values which will give us a measure of thinness and thickness.

an example

- 25 variables and 6 species in our database.

Result of our work

- Our database of 40 leaves were divided into 6 leaves for testing data and remaining for the training data.
- This method identified the correct species for all the testing data.

Future directions

- We have considered a subjective cut-off (for dendogram) on the basis of our data. Further analysis can be done to modify it.
- There were only 6 species in our database. Work can be done further by increasing the data base.