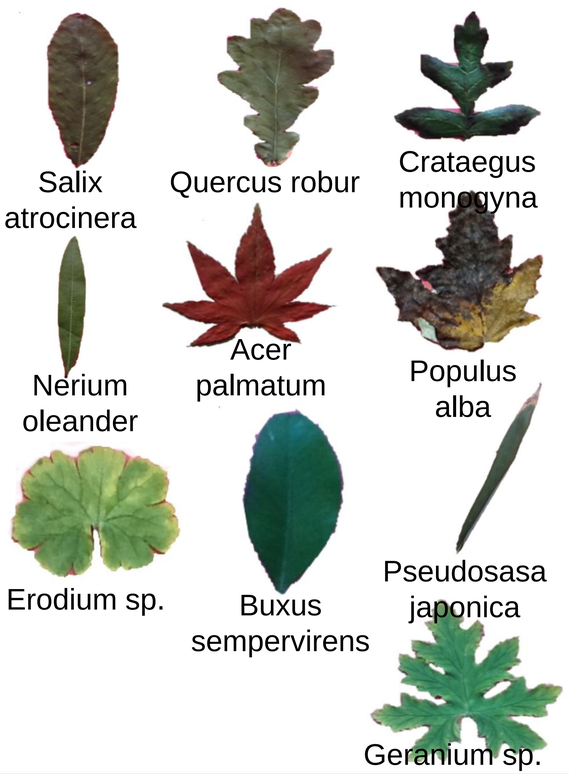
Leaf Classification
Linear Discriminant Analysis

Veniamin Smirnov
Mathematics & Statistics

There are estimated to be nearly half a million species of plant in the world. Classification of species has been historically problematic and often results in duplicate identifications. Automating plant recognition might have many applications, including: Species population tracking and preservation, Plant-based medicinal research, Crop and food supply management.

The objective of this project is to use binary leaf images and extracted features, including shape, margin & texture, to accurately identify 10 species of plants. Leaves, due to their volume, prevalence, and unique characteristics, are an effective means of differentiating plant species.

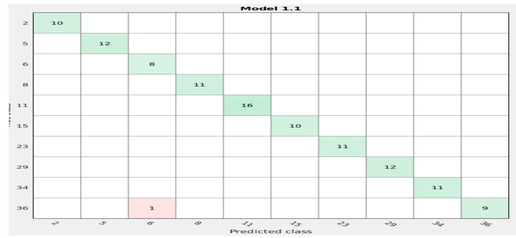


Linear Discriminant Analysis (LDA) is most commonly used as dimensionality reduction technique in the per-processing step for pattern-classification and machine learning applications. The goal is to project a data set onto a lower-dimensional space with good class-separability in order avoid over fitting (“curse of dimensionality”) and also reduce computational costs.

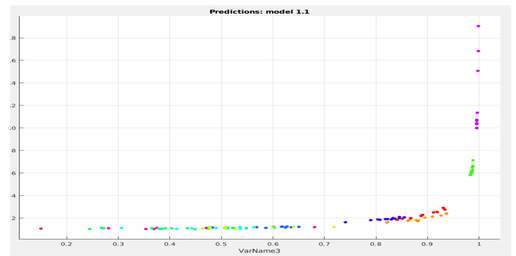
Ronald A. Fisher formulated the Linear Discriminant in 1936 (The Use of Multiple Measurements in Taxonomic Problems), and it also has some practical uses as classifier.

Algorithm

1. Compute the d-dimensional mean vectors for the different classes from the data set.
2. Compute the scatter matrices (in-between-class and within-class scatter matrix).
3. Compute the eigenvectors and corresponding eigenvalues for the scatter matrices.
4. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a d×k dimensional matrix.
5. Use this eigenvector matrix to transform the samples onto the new subspace.



Confussion Matrix. Accuracy is 99.1%



Prediction results by the model

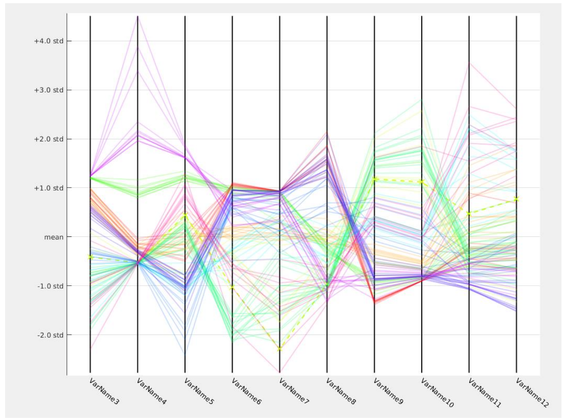
The present project comprises 10 different plant species exhibiting simple leaves. Each leaf specimen was photographed over a colored background using an Apple iPad 2 device.

The following characteristics of the leaves were used:
Eccentricity, Aspect Ratio, Elongation, Solidity, Stochastic Convexity, Isoperimetric Factor, Maximal Indentation Depth, Lobedness, Average Contrast.

Conclusions

With the help of Maple many different algorithms were using along with Linear Discriminant Analysis.

Since the data was heterogeneous some of the algorithms miserably failed, for instance, the accuracy of coarse K nearest neighbor algorithm was about 14.1%, or coarse trees that used Gini’s diversity index provided us with results whose accuracy barely reached 45%, boost trees were very unreliable as well giving us 14% of correct answers.



Future work

Linear Discriminant Analysis can be applied to various problems, such as, face recognition (linear discriminant analysis is primarily used here to reduce the number of features to a more manageable number before classification.