Classification vs Clustering

The process of classifying and clustering appear similar on the surface. In both methods, objects are grouped based on one or more features. Classification is a supervised learning method however, where the labels put on an object are pre-defined and based off that objects features. Labels are provided to the model in this method, alongside some training. Clustering on the other hand, is unsupervised. In clustering, the model groups objects based on object characteristics without prior training and an unknown training sample is provided.

Linear Discrimination Analysis (LDA)

At the most basic level, Linear Discrimination Analysis is used as a dimension reduction tool. By finding correlations between the dimensions in data, LDA helps seek out and drop unnecessary variables in the data. This is an important step in the data cleaning process and optimises training without risking losing crucial information.

The uses of LDA go even further than dimensionality reduction however. It is also a reliable classification tool. It is frequently implemented in facial recognition technology. The main limitation (and benefit) of LDA is its simplicity, however there are extensions to LDA that allow for more flexible decision boundaries.

How does it work?

It focuses on maximising separability between established classes. It can do this by creating a new axis from two existing variables and projects this data onto one axis in two distinct categories. There are two important fundamental statistics required – category means and category variance. The difference between category means should be as far away from each other as possible. LDA aims to reduce the variance within categories. When the distance between means and the variance are optimised, the separation of categories is best. To classify data LDA finds the highest probability for each category.

Example: If the probability of having a fracture with a bone density of 0.54 is greater than the probability of not having a fracture with a bone density of 0.54, a new observation Y is more likely to fall into the category of having a fracture.

Bayes Theory estimates the probability of an output class (k) of an input (x) using the probability of each class and the data to which each class belongs.

References:

<https://machinelearningmastery.com/linear-discriminant-analysis-for-machine-learning/>

<https://youtu.be/D4C7YbfFQSk>