Wine classification

Problem Statement

The data set are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

Classify the types of wines using linear discriminant analysis.

Data Description

The following features are the results of a chemical analysis of three types of wines.

Type

This column describes an observation which belongs to a class.

There are three types of wines.

Alcohol

This attribute refers to the percent alcohol content of the wine (% of volume).

Malic acid

This attribute refers to the percent Malic acid content of the wine (g/L).

Malic acid is one of the primary contributors of acidity in the grape. Its concentration tends to decrease as the grape ripens, mostly due to metabolic respiration. The vine and grape will use malic acid as fuel in respiration.

IMARTICUS LEARNING

Malic acid is very important in wine. If there is not enough, the wine will taste “flat,” and will be more susceptible to spoilage. If there is too much, the wine will taste “green,” or “sour.” Thus it is important for the winemaker to control the amount of Malic acid present.

Ash

Ash content is one of the important indicators in wine quality determination. The conventional gravimetric method for ash quantity determination is time consuming.

Alcalinity of ash

Alkalinity of ash measures the basicity (alkalinity) of the ash obtained from a sample. This is accomplished by adding acid to the ash until the solution is neutralized.

Magnesium

This attribute refers to the percent Magnesium content of the wine (mg).

Magnesium plays a role in the metabolism of carbohydrates and it's believed to be involved in the way insulin is released and the way it behaves in general. Magnesium's role in diabetes is also being studied and it seems there is a link between diabetes and a magnesium deficiency.

Magnesium helps keep the blood's pH levels in balance by controlling the amount of acid in the blood.

Total phenols

The phenolic content in wine refers to the phenolic compounds—natural phenol and polyphenols—in wine, which include a large group of several hundred chemical compounds that affect the taste, color and mouth feel of wine.

These compounds include phenolic acids, stilbenoids, flavonols, dihydroflavonols, anthocyanins, flavanol monomers (catechins) and flavanol polymers (proanthocyanidins).

Flavonoids

Flavonoids mainly derived from the stems, seeds and skins are often leached out of the grape during the maceration period of winemaking. The amount of phenols leached is known as extraction.

These compounds contribute to the astringency, color, and mouth feel of the wine.

In white wines, the number of flavonoids is reduced due to the lesser contact with the skins that they receive during winemaking.

Nonflavanoid phenols

Hydroxycinnamic acids are the most important group of non-flavonoid phenols in wine. The four most abundant ones are the tartaric acid esters trans-caftaric, cis- and trans-coutaric, and trans-fertaric acids. In wine they are present also in the free form (trans-caffeic, trans-p-coumaric, and trans-ferulic acids).

Proanthocyanins

Proanthocyanidins play an important role in wine; with the capability to bind salivary proteins, these condensed tannins strongly influence the perceived astringency of the wine.

Color intensity

The intensity of a wine's color is directly proportional to the intensity of the wine's flavor. Some wines are generally milder while deep rich colors represent more robust flavors.

Hue

A different value of hue and color of wine give different color wines.

Dilution

This attribute refers to the percent dilution content of the wine (g/L).

Proline

Proline is typically the most abundant amino acid present in grape juice and wine. The amount present is influenced by viticulture and winemaking factors and can be of diagnostic importance.

Evaluation Parameters:

Evaluation will be based on:

- Model comparison
- Model selection

Model comparison

Apply LDA algorithms and compare results

Model selection

Select the best model. Model selection to be based on Accuracy, Sensitivity, Specificity, Positive Predictive value, Negative Predictive Value, Prevalence, Detection Rate, Detection Prevalence and Balanced Accuracy.

Expected Outcome:

Higher accuracy in predicting the outcome using test data.