Predicting and classifying wines based on physical and chemical properties

January 6, 2019

Something something wine is very important, and somewhat unique in the modern age; whilst most consumer goods are specified and produced in a controlled manner with six sigma type methods, wine varies significantly not just between brands, but between batches.

The question set is:

"Chemically speaking, what types of wine are there? What predicts wine quality?"

This question comes in two parts: the latter is more traditionally suited to regression techniques, the former appears little more towards neural net approaches, although testing for collinearity and bimodals can still be of use.

The dataset consists of 6497 wines (1599 red, 4898 white) which have been assigned a quality rating (the dependent variable) from 0-10. 10 (independent) variables have been measured: fixed and volatile acidity, citric acid levels, residual sugar, chlorides, free and total sulfur dioxide, density, pH, sulphates and alcohol content (the dataset is complete, though units are not provided).

On a superficial level, there are clearly two types of wine: red and white, although that may be considered anthropomorphic. T-test of the two datasets confirms they are indeed different in every measured variable except alcohol content, with red wines having greater acidity, less residual sugar, and less sulfur dioxide.
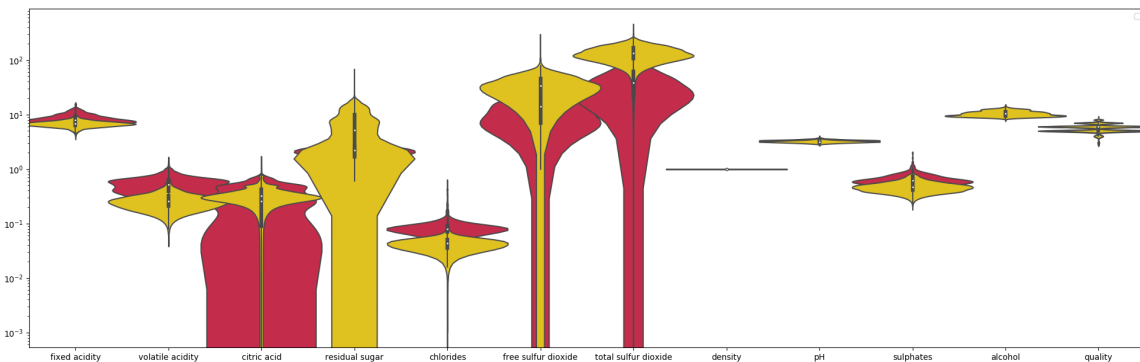

Figure 1: mean properties for red and white wines

Finding predictors

An initial hypothesis would be that winemakers have roughly optimised each chemical present, so that the optimum lies somewhere within the space tested: in this case a second-degree polynomial (with second-order term negative) is appropriate. Equally possible is that some quantities should be strictly minimised or maximised, i.e. linear.

Properties matching the stricter quadratic criteria are: for red wines, sulphates and citric acid; for white wines, free and total sulfur dioxide. For red wine only, increasing alcohol content increased perceived quality ($\alpha < 0.01$). For both types, volatile acidity, chloride and density should be minimised.

	red	white
fixed acidity		
volatile acidity	↓	↓
citric acid		0.553
residual sugar		↓
chloride	↓	↓
free sulfur dioxide		0.00364
total sulfur dioxide	↓	0.00206
density	↓	↓
pH		
sulphates	0.385	↓
alcohol	↑	

Figure 2: ideal properties for red and white wines; arrows indicate no optimum was found