Proposal

Qi Wang

2021/10/26

1. Introduction

2. Our Goal of Research

Here, in this paper, we are going to explore some methods to classify the wine into different quality levels. We do not use clustering methods here but we will use logistic regression to achieve our goal. And our goal is to explore which of these variables which I will talk about in the variable description section will have a significant effect on the red wine qualification and it is positive or negative, and how much will the significant variables affect the level of the red wine qualification.

3. Variable Despription

There are 11 covariates and one categorical response variable. Fix acidity is the most acids involved with wine or fixed or nonvolatile (do not evaporate readily). Volatile acidity is the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste. Citric acid is the most important organic acid produced in tonnage by fermentation, with a taste of sour like lemons (Soccol et al. 2006). Residual sugar is the amount of sugar left in a wine, to some extent, it measures the sweetness of a wine. Chlorides is a key role in the salty taste of a wine, which will make customers feel uncomfortable. Sulfur dioxide (SO_2) is important in the winemaking process as it aids in preventing microbial growth and the oxidation of wine (Monro et al. 2012). The difference between free sulfur dioxide and total sulfur dioxide is the way of measuring them. Gaseous SO2 is released from the sample by addition of acid and swept into the ICP by an argon stream. The intensity of the sulfur atomic emission lines is measured in the vacuum UV region. Determination of total SO_2 is performed after hydrolysis of bound forms with sodium hydroxide ($NaOH$)(Čmelík et al. 2005). For sulfates, many experts believe that higher sulfurous content causes a duller taste in wine, and that high potency of sulfite ions presents a health risk and speeds up the wine's fermentation process. The other covariates including alcohol, pH and density are basically simple indexes of a red wine. Our response variable is an ordered categorical variable indicating the quality of red wine, from 0 to 10.

4. Problems When Analyzing

1. To begin with, the response variable in this data set is categorical but have 11 degrees, although in the data there are only five levels, we still need to first merge some groups to get a 0-1 variable to continue the logistic regression.

2. Then, there are some inner colinearship among the covariates. Citric acid amount is strongly positively correlated with fixed acidity. And fixed acidity is also strongly positively correlated with the density

of the wine. Researches have shown that the citric acid has an effect on the acidity of the liquid, and a more concentration of citric acid means a stronger acidity (Lustig et al. 2017).Also, the amount of free sulfur dioxide is positively correlated with the amount of total sulfur dioxide. However, there are still some negative correlations among variables. For example, the pH of the red wine is negatively correlated with the fixed acidity and the amount of citric acid in the wine. From common sense, we know a stronger acidity means a lower pH, that's why the pH is lower for those wine with more concentration of fixed acidity and citric acid. Also, larger concentration of alcohol gives a smaller density. As we know, the density of alcohol is smaller than water, so if more alcohol is included in the wine, the density must be lower than those without that much alcohol. I fixed this problem by first select variables for a base model which has weak relationships, then add variables to check the model selection criteria.

3. There are some outliers in the model, but after checking the resource of data, I don't think it is a measurement error, therefore, I just leave it there and still looking for further methods to handle it.

4. It is a logistic regression, so the residual analysis is tricky. I will begin this part after learning this method in the class.

References

Čmelík, Jiří, Jiří Machát, Eva Niedobová, Vítězslav Otruba, and Viktor Kanický. 2005. "Determination of Free and Total Sulfur Dioxide in Wine Samples by Vapour-Generation Inductively Coupled Plasma–Optical-Emission Spectrometry." *Analytical and Bioanalytical Chemistry* 383 (3): 483–88.

Lustig, William P, Soumya Mukherjee, Nathan D Rudd, Aamod V Desai, Jing Li, and Sujit K Ghosh. 2017. "Metal–Organic Frameworks: Functional Luminescent and Photonic Materials for Sensing Applications." *Chemical Society Reviews* 46 (11): 3242–85.

Monro, Tanya M, Rachel L Moore, Mai-Chi Nguyen, Heike Ebendorff-Heidepriem, George K Skouroumounis, Gordon M Elsey, and Dennis K Taylor. 2012. "Sensing Free Sulfur Dioxide in Wine." *Sensors* 12 (8): 10759–73.

Soccol, Carlos R, Luciana PS Vandenberghe, Cristine Rodrigues, and Ashok Pandey. 2006. "New Perspectives for Citric Acid Production and Application." *Food Technology & Biotechnology* 44 (2).