Red Wine Review

Using Logistic Regression, LDA and Nonparameteric model approach

MSDS6372 – Project 2

November 2018

Vitaly Briker | Kelly Carter | An Nguyen

Table of Contents

[Introduction 3](#_Toc530320891)

[Data Description 3](#_Toc530320892)

[Exploratory Analysis 3](#_Toc530320893)

[Objective 1 – Logistic Regression Model 3](#_Toc530320894)

[Problem Statement and Approach 3](#_Toc530320895)

[Assumptions 3](#_Toc530320896)

[Model Fit 3](#_Toc530320897)

[Parameter Interpretation 3](#_Toc530320898)

[Conclusion 3](#_Toc530320899)

[Objective 2 – Logistic Regression 2 4](#_Toc530320900)

[Summary 4](#_Toc530320901)

[Model Fit 4](#_Toc530320902)

[Parameter Interpretation 4](#_Toc530320903)

[Conclusion 5](#_Toc530320904)

[Objective 2 – LDA 5](#_Toc530320905)

[Problem Statement and Approach 5](#_Toc530320906)

[Assumptions 5](#_Toc530320907)

[Model Fit 6](#_Toc530320908)

[Parameter Interpretation 6](#_Toc530320909)

[Conclusion 6](#_Toc530320910)

[Objective 2 – Nonparametric approach 6](#_Toc530320911)

[Problem Statement and Approach 6](#_Toc530320912)

[Assumptions 6](#_Toc530320913)

[Model Fit 6](#_Toc530320914)

[Parameter Interpretation 6](#_Toc530320915)

[Conclusion 6](#_Toc530320916)

# Introduction

In this project, we attempted to classify red wines that are related to the Portugese “Vinho Verde” variants. Using a data set provided by the UCI machine learning repository, <https://archive.ics.uci.edu/ml/datasets/wine+quality>. The data set included fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, and quality as variables. Using the quality variable, based on sensory data score between 0 and 10, we categorized the wines into a ‘poor’ category or a ‘fine’ category; 5 or less in quality was desiganted as ‘poor’ and greater than 5 is ‘fine’. For the purpose of this project our objectives are as follows:

1. Build a logistic regression model using the provided data.
2. Building on the regression model above add complexity to the model.
3. Create a competing model using LDA or QDA.
4. Use a nonparametric model approach in a third model.

## Data Description

Source: https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009

* Contains 1599 records

For the purpose of this project, we randomly split the data set into a training set and a test set 50/50:

* train: wine\_train.csv
* test: wine\_test.csv

## Exploratory Analysis

Kelly

# Objective 1 – Logistic Regression Model

## Problem Statement and Approach

Kelly

## Assumptions

Kelly

## Model Fit

Kelly

## Parameter Interpretation

Kelly

## Conclusion

Kelly

# Objective 2 – Logistic Regression 2

## Summary

While running the logistic regression model using the provided data it was noticed that some of the variables were not normally distributed and were very skewed, namely residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, sulphates, alcohol and fixed acidity. A log transformation was applied to all the named variables except for the fixed acidit as it had 0’s.

## Model Fit

After applying the Forward selection method to the log transformed and non-transformed variables seven variables were selected by the algorithm. The Hosmer-Lemshow test concluded that the model had a good fit (p = 0.7738). There was no noticebly large outliers or leverage points, diagnostic table can be found in the appendix. The AIC model fit statistics is 825.142. SAS output can be found in the appendix

## Parameter Interpretation

Log(odds) = -11.7 – 1.1(log(chlorides)) – 0.3(log(total sulfur dioxide)) + 3.6(log(sulphates) + 9.0(log(alcohol)) – 3.2(volatile acidity) – 2.2(citric acid) – 2.0(pH)

## Conclusion

# Objective 2 – LDA

## Problem Statement and Approach

AN

## Assumptions

AN

## Model Fit

AN

## Parameter Interpretation

AN

## Conclusion

AN

# Objective 2 – Nonparametric approach

## Problem Statement and Approach

Vitaly

## Assumptions

Vitaly

## Model Fit

Vitaly

## Parameter Interpretation

Vitaly

## Conclusion

Vitaly