

Linear Regression



0.5706



0.5663



0.5771

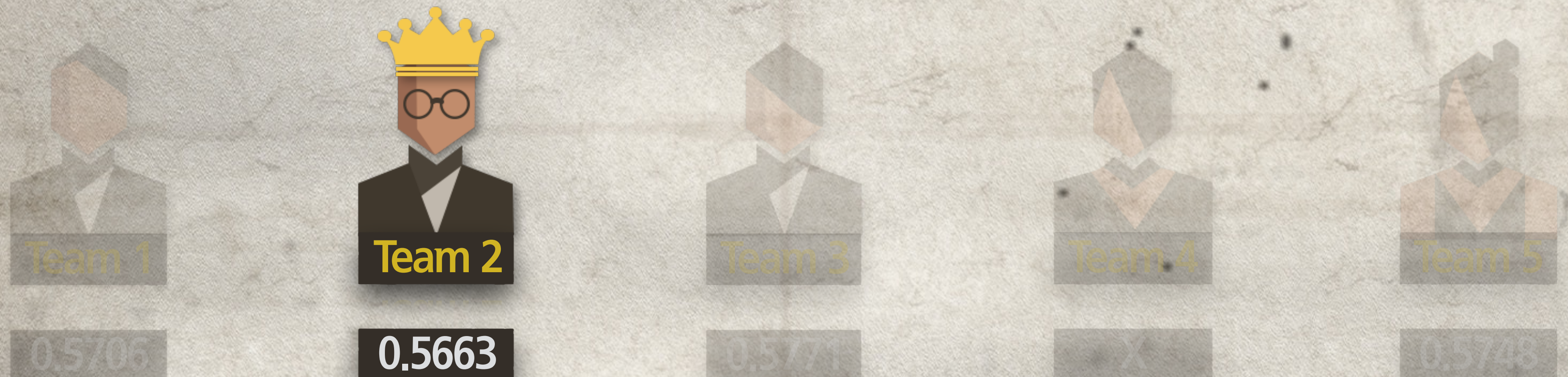


X



0.5748

Linear Regression



Predicting Wine Quality

with
Linear Regression

20160802

Han Jaeyoon / Dept. of SNS



Main Idea



Bitter

Less Sugar

Dry

Fruity, Not Bitter

More Sugar

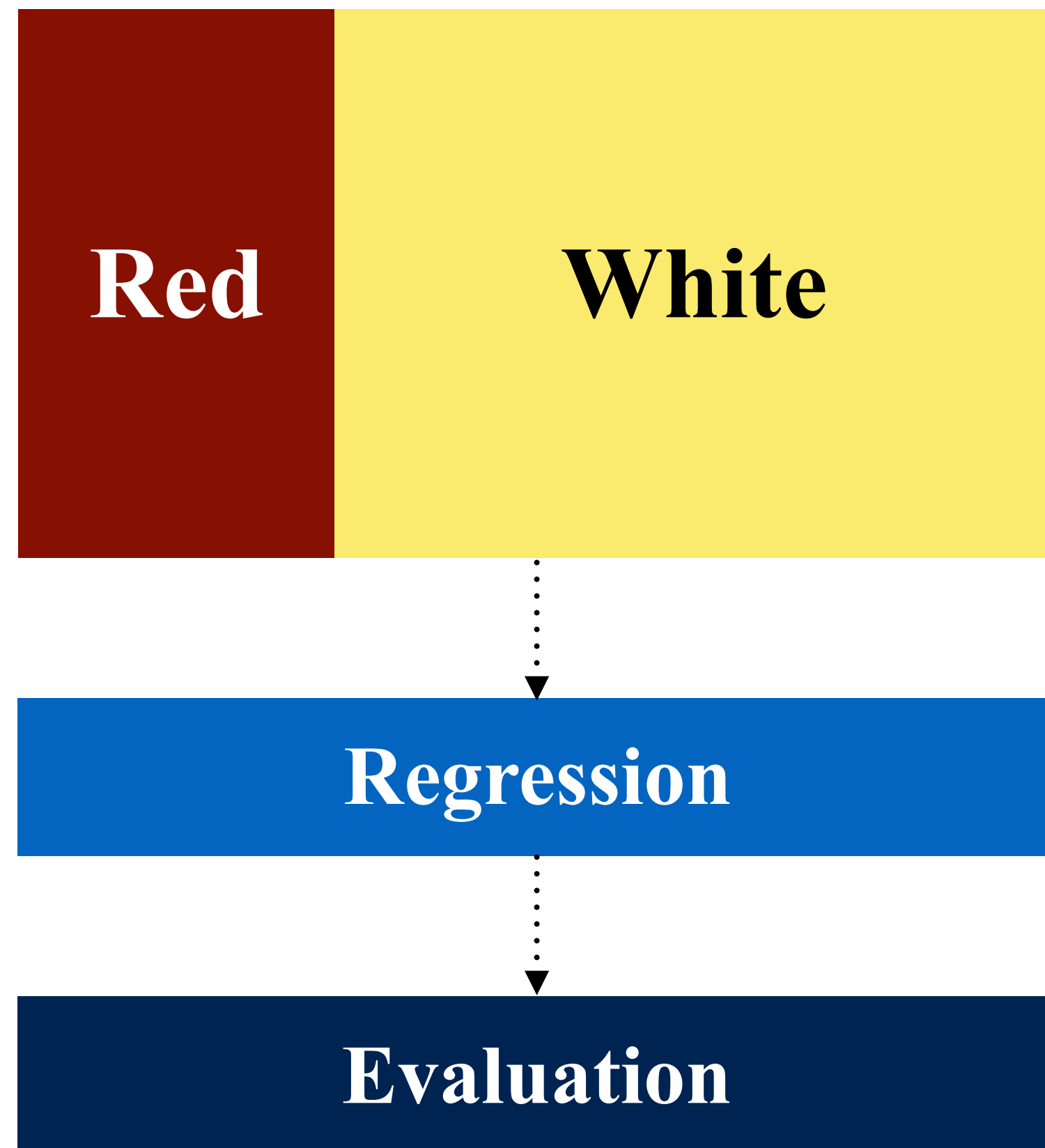
With Dessert



There are **different important factors** between two types of wines.

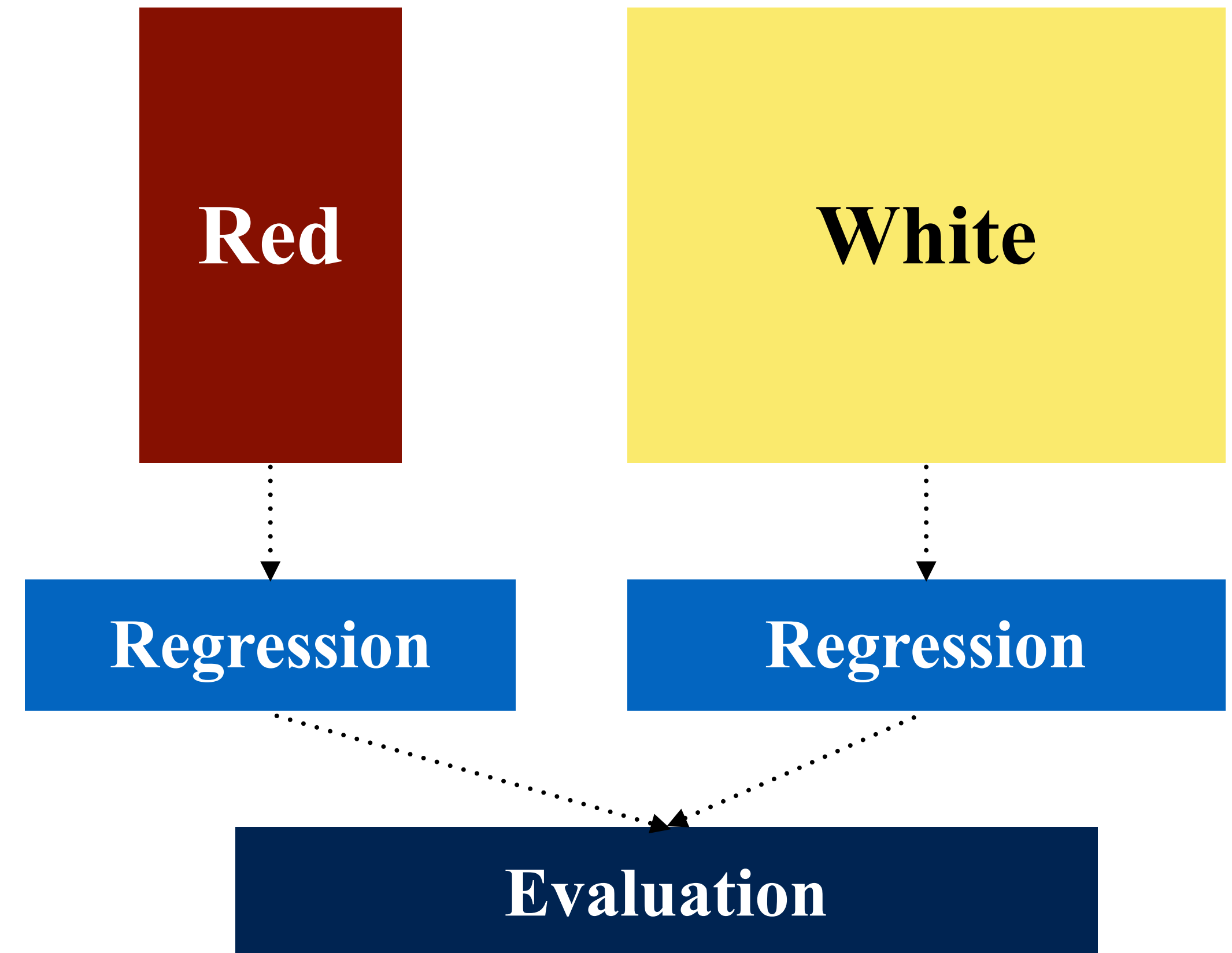
Main Idea

Traditional Method



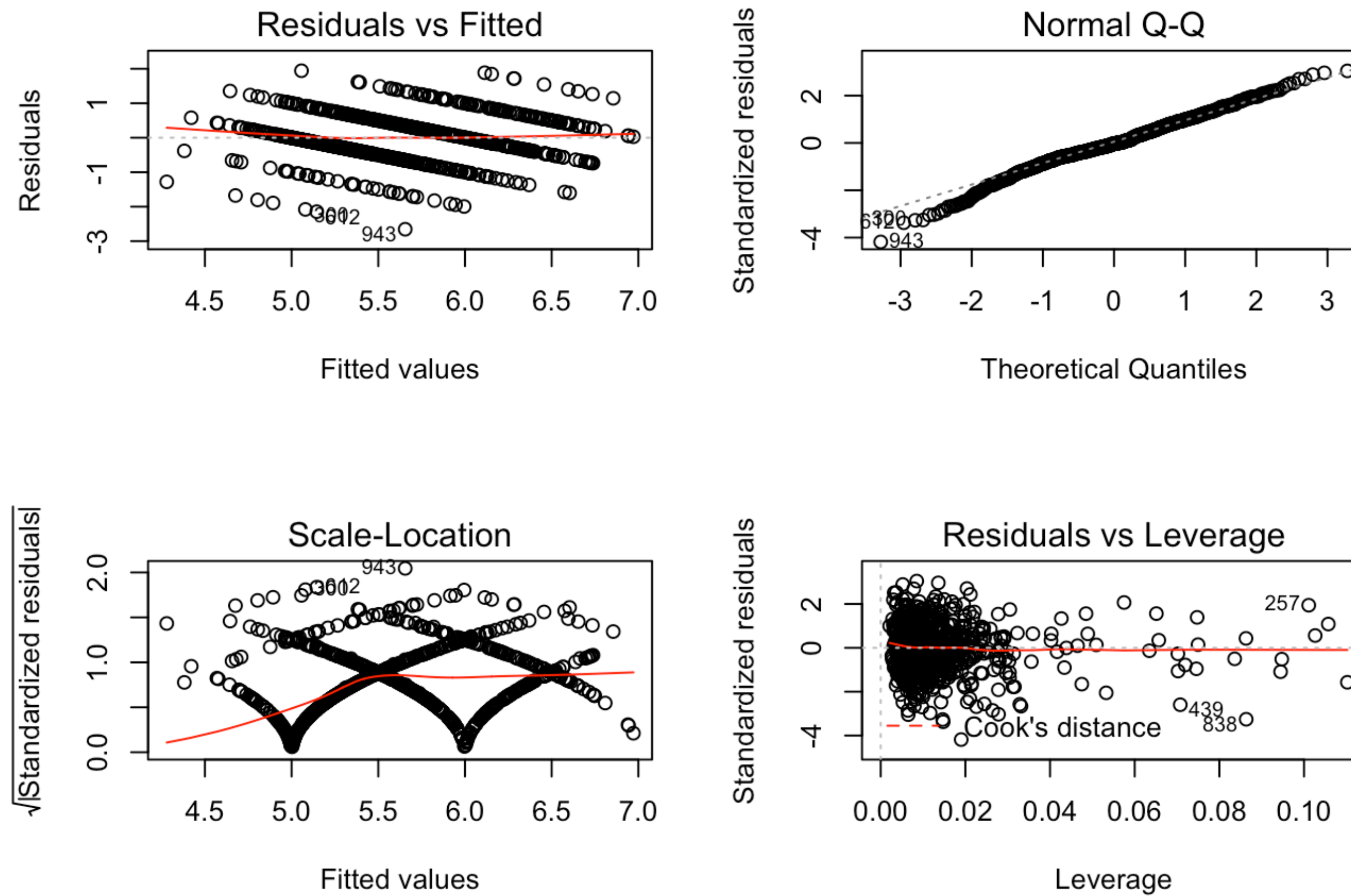
The Problem happens with Feature Selection

My Method



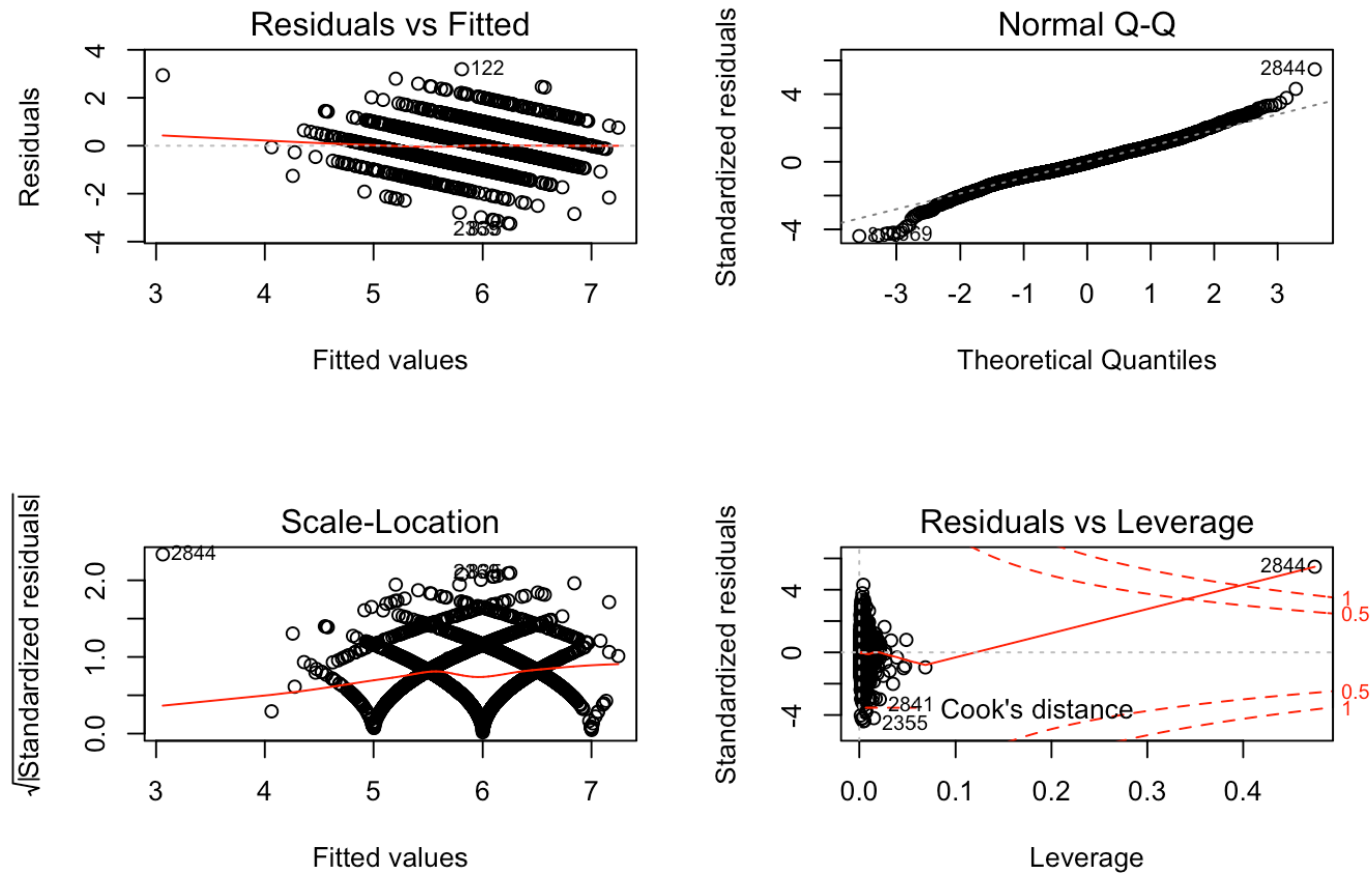
It can be considered as an ensemble model
with two independent models

Outlier & Leverage Detection



Default : Red Wine

Outlier & Leverage Detection



Default : White Wine

Feature Selection & Interaction Term

Attributes Description

1. fixed acidity: most acids involved with wine or fixed or nonvolatile (do not evaporate readily)
2. volatile acidity: the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste
3. citric acid: found in small quantities, citric acid can add 'freshness' and flavor to wines
4. residual sugar: the amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and wines with greater than 45 grams/liter are considered sweet
5. chlorides: the amount of salt in the wine
6. free sulfur dioxide: the free form of SO_2 exists in equilibrium between molecular SO_2 (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine
7. total sulfur dioxide: amount of free and bound forms of SO_2 ; in low concentrations, SO_2 is mostly undetectable in wine, but at free SO_2 concentrations over 50 ppm, SO_2 becomes evident in the nose and taste of wine
8. density: the density of water is close to that of water depending on the percent alcohol and sugar content
9. pH: describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale
10. sulphates: a wine additive which can contribute to sulfur dioxide gas (SO_2) levels, which acts as an antimicrobial and antioxidant
11. alcohol: the percent alcohol content of the wine
12. type : the type of the wine

Feature Selection & Interaction Term

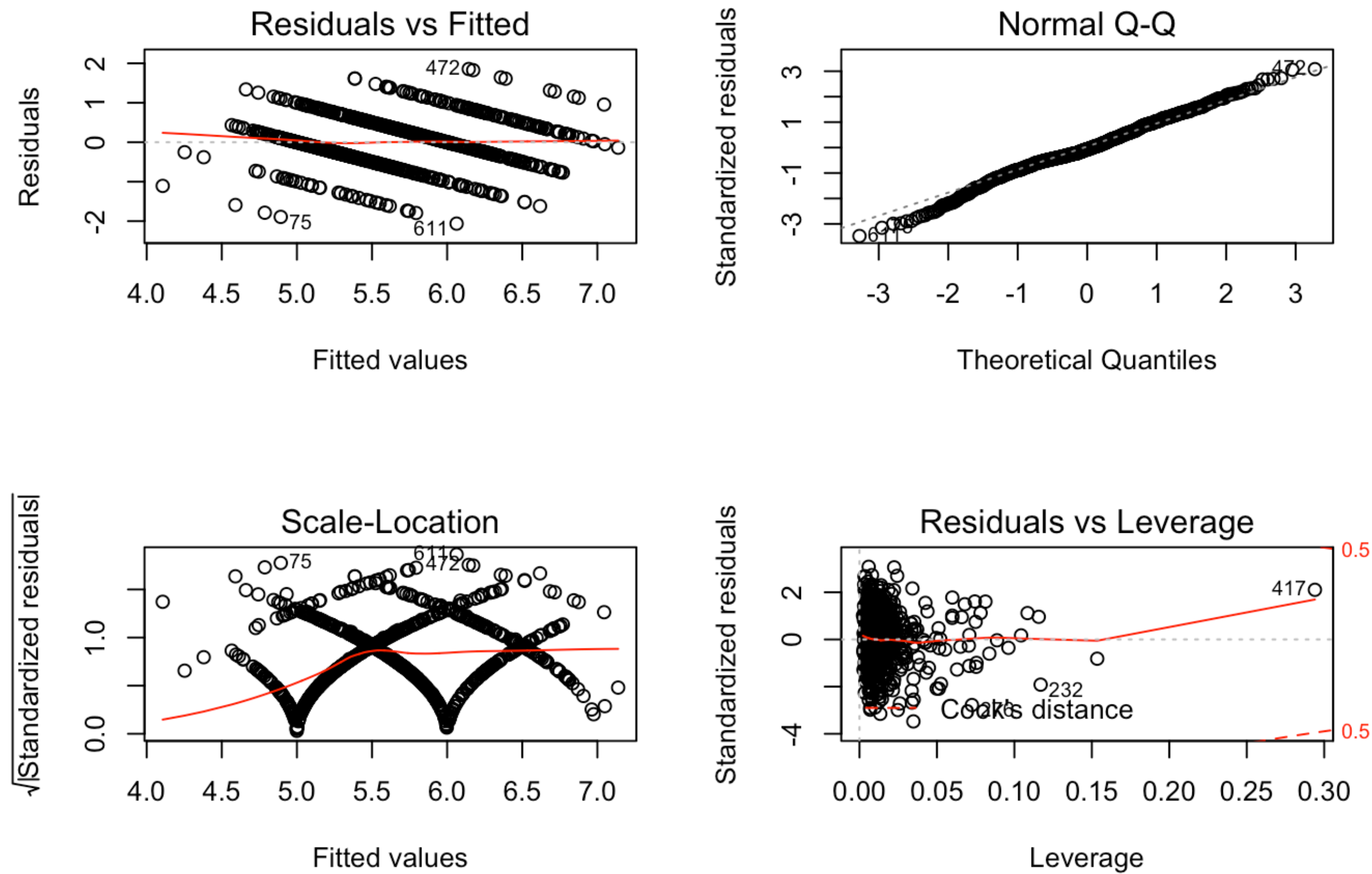
For Red Wine

```
step_red <- lm(formula = quality ~ volatile.acidity + citric.acid +  
               residual.sugar + chlorides + free.sulfur.dioxide * density *  
               total.sulfur.dioxide * sulphates + pH + density * alcohol,  
               data = new_red)
```

For White Wine

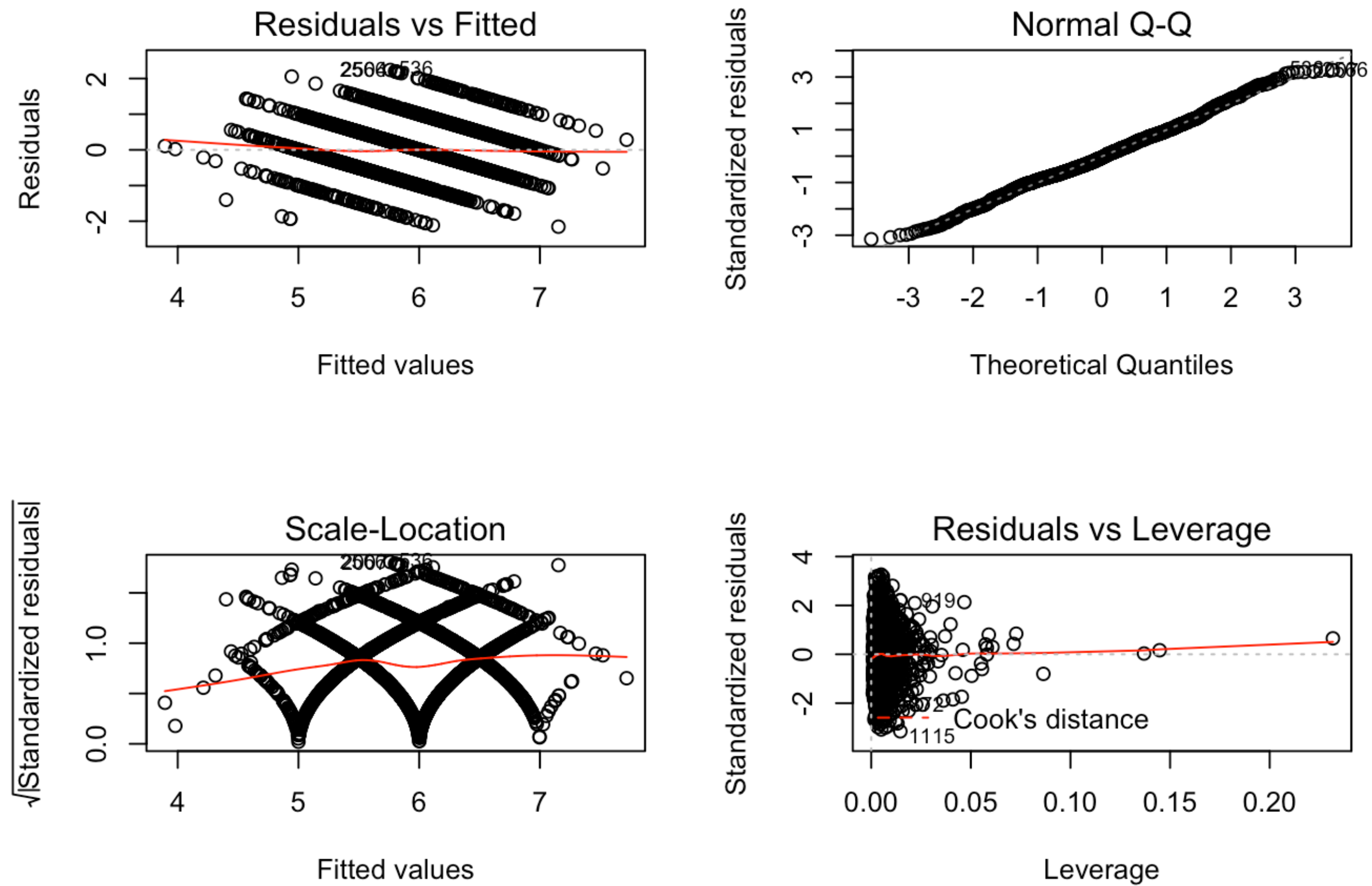
```
step_white <- lm(formula = quality ~ fixed.acidity * volatile.acidity +  
               residual.sugar * pH + free.sulfur.dioxide * total.sulfur.dioxide *  
               sulphates + density * alcohol, data = new_white)
```


Outlier & Leverage Detection



Featured Model : Red Wine

Outlier & Leverage Detection



Featured Model : White Wine

Result

Model	MAE
Traditional Model	0.5814709
Ensemble Model	0.5769779
Ensemble Model + Outlier & Leverage Detection (1)	0.5779364
Ensemble Model + Outlier & Leverage Detection (1) + Interaction Term	0.5657421
<i>Ensemble Model + Outlier & Leverage Detection (1) Interaction Term + Outlier & Leverage Detection (2)</i>	<i>0.5641923</i>

Question?