

Assignment 1

Overview of data

PCA analysis

1. Plots of principle components
2. Comment on the main chemical differences between red and white wines
3. Including quality variable what features are likely to be present in wines of good quality? Is it different for red and white?

If space allows - further EDA

Perform visualisations - any trends between variables? Something interesting if datasets are considered separately?

One page - Hotelling T square test

1. Perform a Hotelling's T^2 -test to test the hypothesis that the red and while wines have the same acidity means (the variables fixed acidity, volatile acidity and pH)
2. Select some variables and computer μ_W of selected variables for white wine
3. Perform 1-sample T^2 test to check whether corresponding means for red wine dataset are equal to μ_W

Your report should have: 1. A question stated clearly 2. Explaining the statistical test used to address question 3. Give solution 4. Potentially visualisation

Other notes: 1. You must give references 2. Try to keep it to 4 pages¹

References

- [1] P. Cortez et al. "Modeling wine preferences by data mining from physicochemical properties". In: *Decis. Support Syst.* 47 (2009), pp. 547–553.

¹Cortez et al. [1]