FINAL REPORT

SIMULATON STUDY REPORT (SINGLE ITERATION PROCEDURE)

A simulation study iteratively executed once on the randomly generated variables from a normal distribution for both the best subset selection and lasso generated the following ME and no.zeros based on the BIC selection Criterion

Method	ME	Number_of_zero_coefficients
Best Subset	0.05279	5
Lasso	0.22754	4

SIMULATION STUDY REPORT (PROCEDURES REPEATED 1000 TIMES)

A simulation study on randomly generated variables from a normal distribution was performed to assess the performance of both best subset selection and Lasso. The procedures were repeated 1000 times, resulting in 1000 values for ME and no.zeros.

Below is the median of the values for each performance measure from the BIC selection Criterion.

Method	ME	Number_of_zero_coefficients
Best Subset	0.14910	5
Lasso	0.38362	4

REPORT ON REAL DATA ANALAYSIS

INTRODUCTION

The aim of the analysis is to report on which variables were chosen from the best subset selection and lasso methods. The dataset under consideration is on quality of Red Wine.

DATA DESCRIPTION

This dataset is related to red variants of the Portuguese "Vinho Verde" wine. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.).

Content

Input variables (based on physicochemical tests):

1 - fixed acidity	2 - volatile acidity	3 - citric acid
4 - residual sugar	5 – chlorides	6 - free sulfur dioxide
7 - total sulfur dioxide	8 – density	9 - pH
10 – sulphates	11 – alcohol	

Output variable (based on sensory data):

12 - quality (score between 0 and 10)

BEST SUBSET SELECTION ANALYSIS ON THE RED WINE QUALITY DATASET

To identify the most influential variables that contribute to the model's predictive power(predicting the quality of red wine), we performed the Best Subset Selection. The following variables were selected

volatile.acidity	chlorides	total.sulfur.dioxide	pH	sulphates	alcohol
-0.2302	-0.1167	-0.0966	-0.0831	0.1865	0.3836

LASSO ANALYSIS ON THE RED WINE QUALITY DATASET

Following the best subset selection, we applied the Lasso method also on the red wine quality dataset. The resulting selected variables, along with their estimated coefficients, are shown below

volatile.acidity	chlorides	free.sulfur.dioxide	total.sulfur.dioxide
-0.2276	-0.0953	0.0241	-0.1009

pH	sulphates	alcohol
-0.0689	0.1688	0.3754