Regression Analysis

Regression coefficients are one of the main features of a regression model. They are the values you use to study the relationship between variables.

1. R-squared

It is a statistical measurement, which is also called the coefficient of determination. It’s a measure of how much of the variation in your independent variable, y, can be explained by variation in the independent variables in your regression. We use R-squared as an indicator of how strong a model is.

1. MSE

It is used to check how close estimates or forecasts are to actual values. Lower the MSE, the closer is forecast to actual. This is used as a [model evaluation measure](https://www.mygreatlearning.com/blog/model-evaluation-techniques-for-machine-learning-classification-models/) for regression models and the lower value indicates a better fit.

1. Regression Coefficients

Each regression coefficient represents the estimated size and direction of the relationship between the dependent variable—also called the response variable—and a particular independent variable—also called a predictor variable.

***Is The Regression Coefficient The Same As The Correlation Coefficient?***

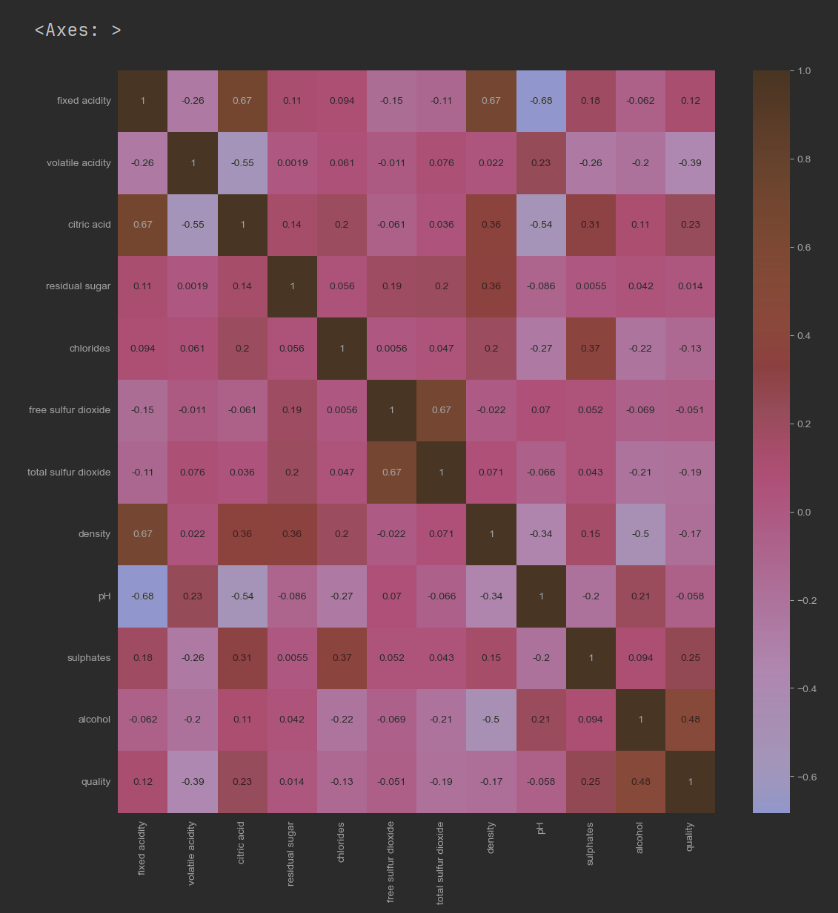
The regression coefficient and the correlation coefficient are two statistical measures we commonly use to evaluate the relationship between two variables, but they are not the same thing.

A **regression coefficient** is an estimate of the change in the dependent variable, Y, that results from a change in the independent variable, X. In other words, it tells us how much Y changes when X changes. In a standard linear regression, the coefficient measures the change in Y that results from a one-unit increase in X.

A **correlation coefficient**, on the other hand, measures the strength of the relationship between two variables. One of the most common correlation coefficients is the Pearson correlation coefficient (r), which measures the strength and direction of the linear relationship between two variables. Unlike a regression coefficient, the correlation coefficient ranges from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation.

1. Correlation

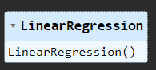
Correlation describes the relationship between variables. It can be described as either strong or weak, and as either positive or negative.

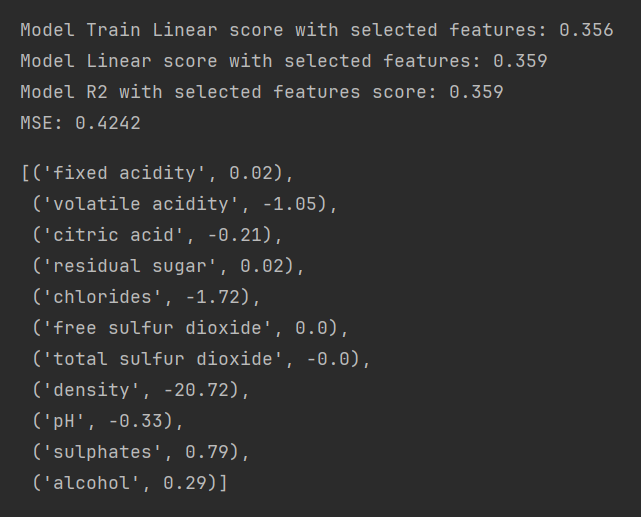


1. Linear Regression

Linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables

* 1. Basic regression





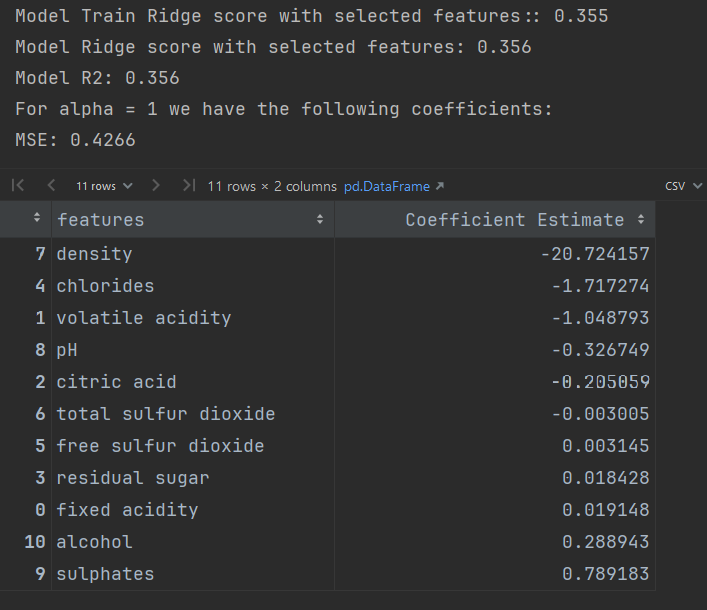
1. Ridge Regression

Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where the independent variables are highly correlated.

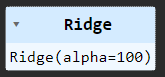


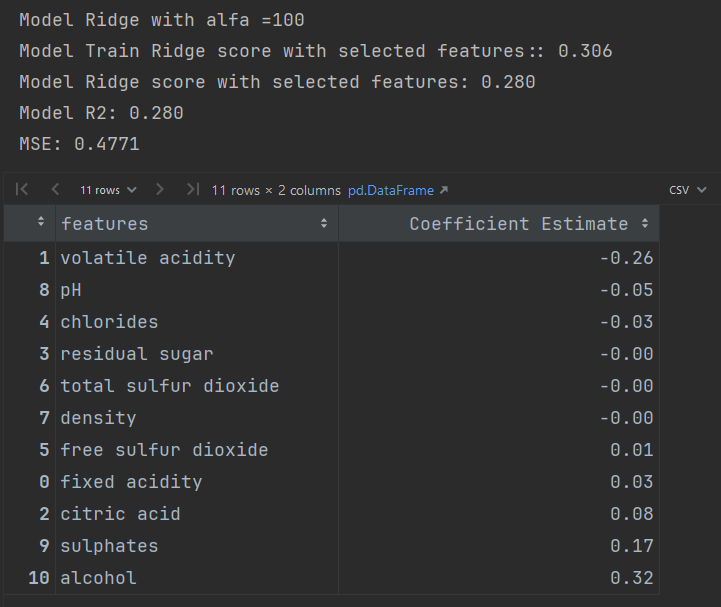
Ridge regressor to extract coefficients

* 1. Alpha = 1

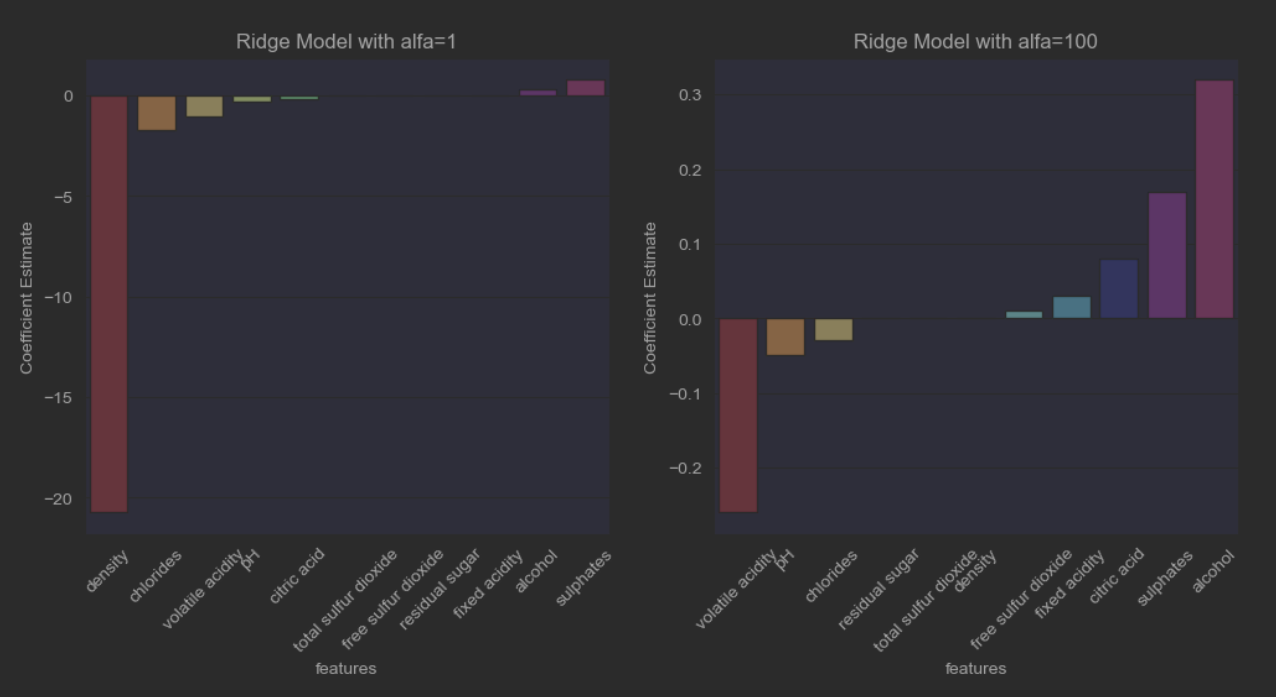


* 1. Alpha = 100



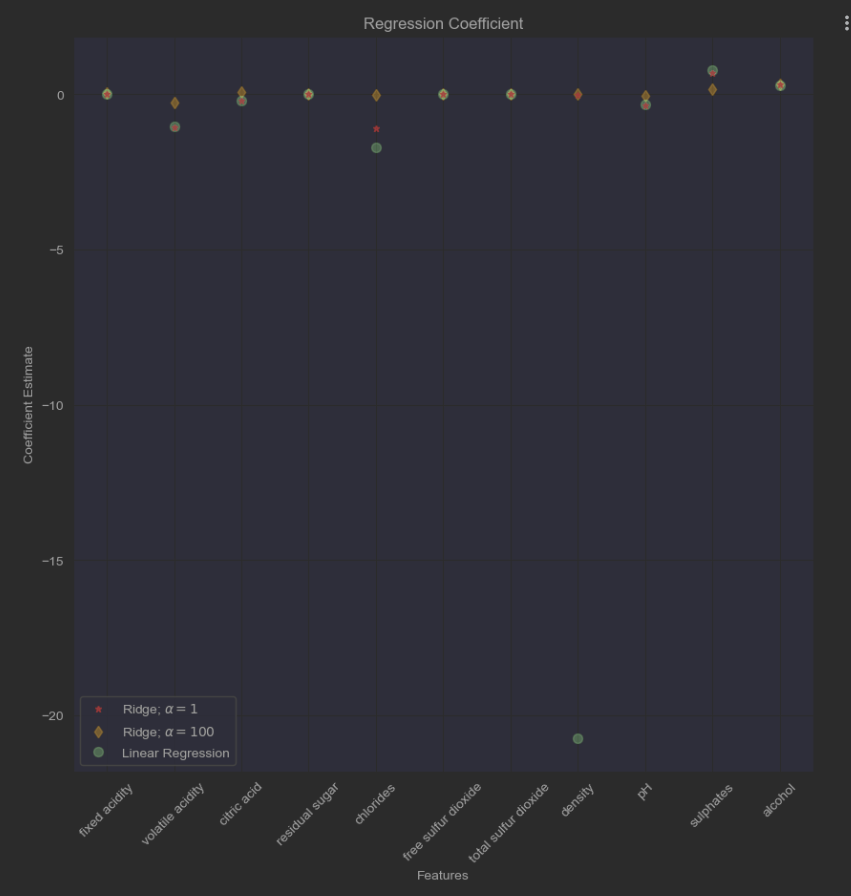


* 1. Alpha effect



The above plot shows that the value of Alpha in Ridge regression influences the direction of the regression coefficients.

* 1. Regression Coefficients of All model features



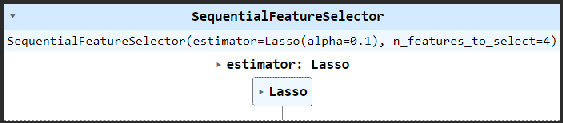
The above plot shows the linear coefficient distribution has much larger variance as we can see some coefficient outliers.

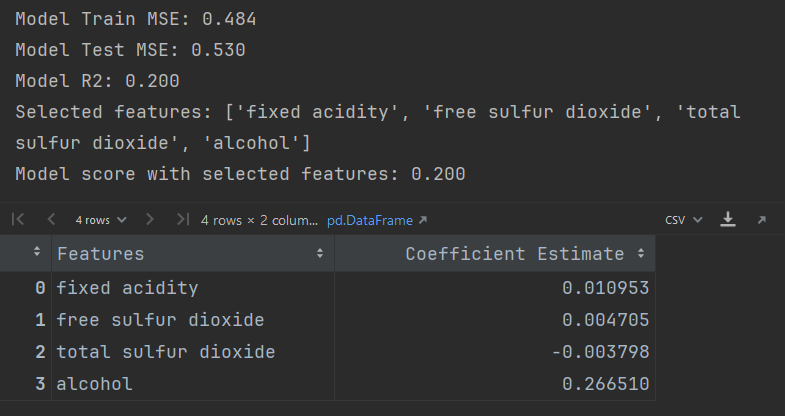
1. Lasso Regressions

LASSO regression, also known as L1 regularization, is a popular technique used in statistical modelling and machine learning to estimate the relationships between variables and make predictions. LASSO stands for Least Absolute Shrinkage and Selection Operator.  
  
Lasso with Sequential Feature Selection

* 1. Lasso with Sequential Feature Selection (four features)

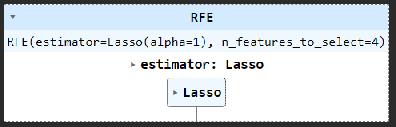
This approach is using sequential feature selection to sequentially add or delete features and only use those that improve the model.

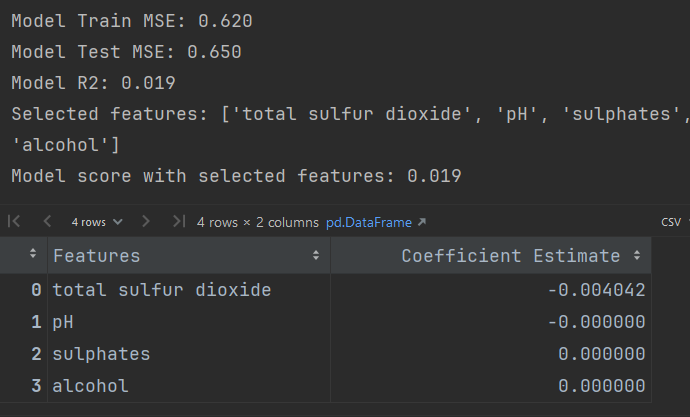




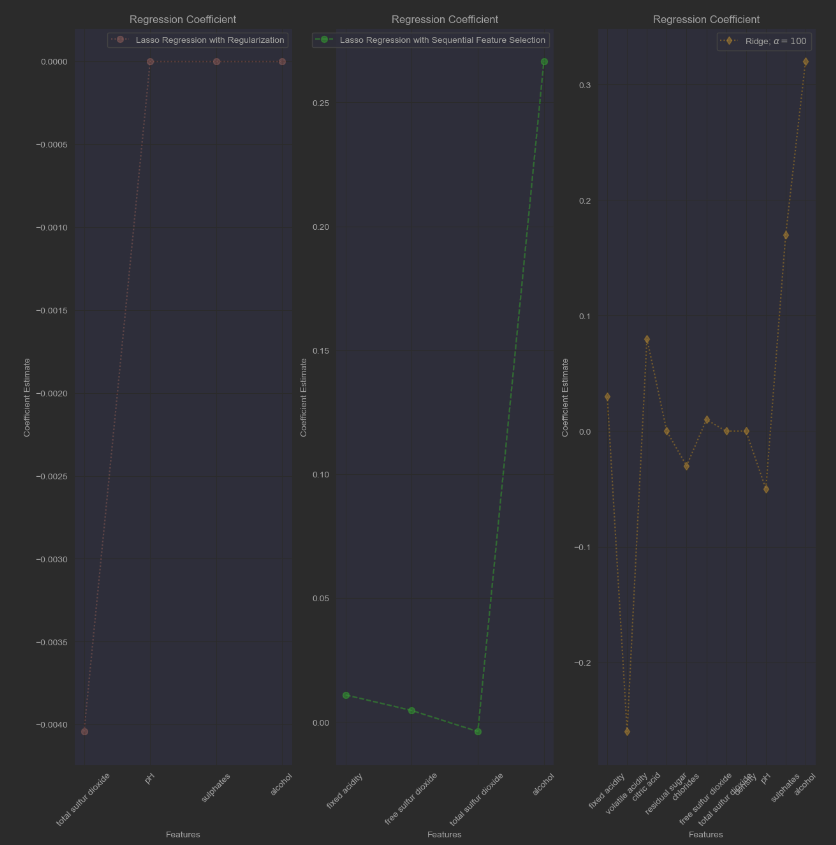
* 1. Lasso Regression with Regularization (four features)

This approach is using a regularized model to identify features based on non-zero coefficients.





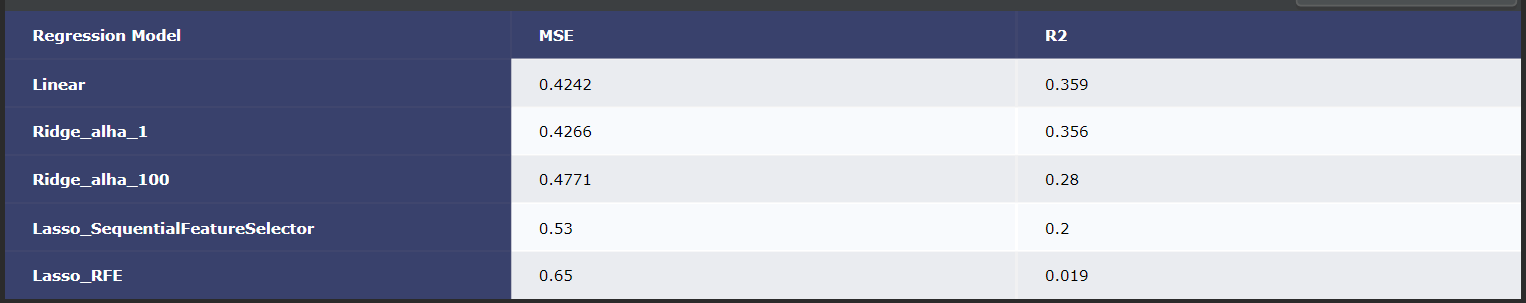
1. Comparison of Regression Coefficient between Lasso and Ridge

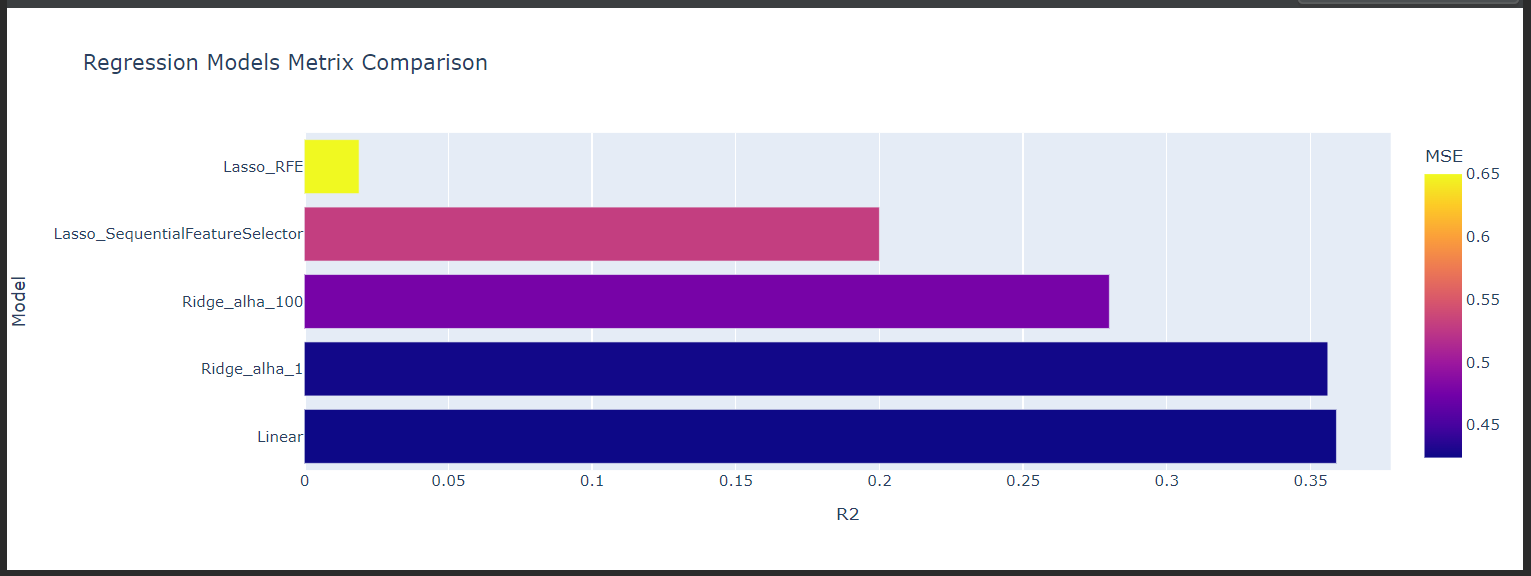


Where are ridge and lasso regression used?

Ridge regression is used in cases where there are many parameters or predictors that affect the outcome. Whereas lasso regression works better when there are fewer significant parameters or predictors involved.

1. Regression Models Comparison





If we look at R2 and MSE scores the Linear and Ridge with alpha=1 are best performers, It looks using the all features create a better model for the above dataset.

Hi Jim,

Good work, I have got very similar values of coefficients, the bar plot is my preference as well, but you cannot see the zero values, so I switched to a different kind of plot.

Hi Radhika,

Good work, the ridge regression approach will always select all features, but you can change the alpha value which affects coefficients, the direction of the relationship between the dependent variable and predictor variable.