

Logistic Regression

Dataset

The dataset used in training the model is called the Wine Quality Dataset from UCI Machine Learning Repositories (source: <https://doi.org/10.24432/C56S3T>). It is a multivariate dataset that can be used in Classification and Regression tasks and is composed of 11 features with 4898 instances. I only selected five instances of this dataset namely 'fixed acidity', 'volatile acidity', 'citric acid', 'sulphates', and 'alcohol' in training the logistic regression model. The data was then scaled between 0 and 1 using the MinMaxScaler and separated into training and testing sets on an 8:2 ratio. The data has the target variable 'quality' with values from 3 to 8 which means that the dataset has multiple classification.

Results

Two regularization method was used in training the Logistic Regression model: the L1 and L2 Regularization. Using the L1 Regularization or Lasso Regression, the results show a significant decrease in accuracy of the model with the accuracy value of 54.69%. Meanwhile, using the L2 Regularization also called the Ridge Regression, the model shows a better performance compared to Lasso Regression with an accuracy value of 55.31%.

