TECHNISCHE HOCHSCHULE INGOLSTADT
ALGORITHMS FOR AI 3

LAB 2: GAUSSIAN PROCESS CLASSIFIER

FERNANDO FINCATTI

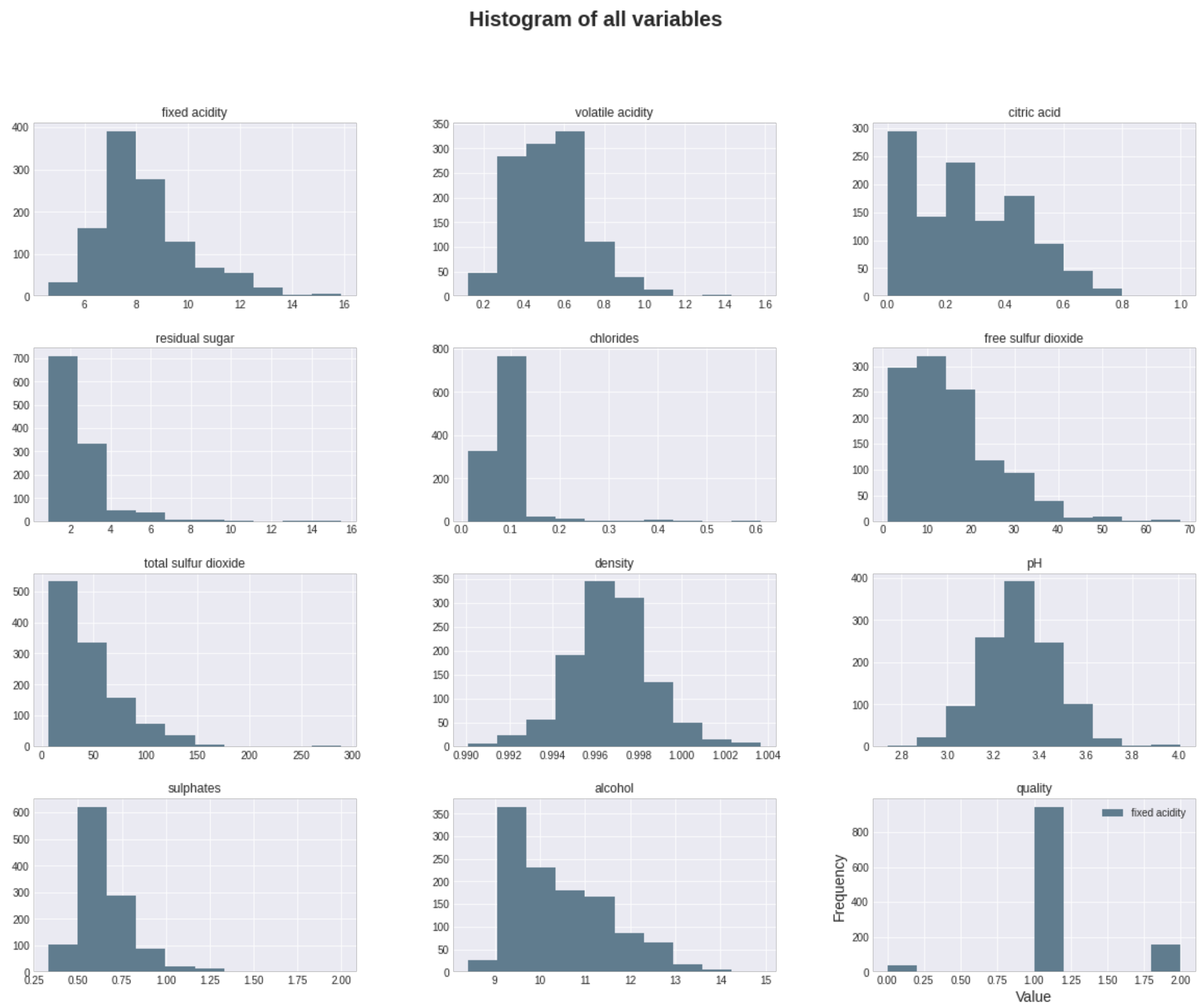
INGOLSTADT
MAY 2023

1. INTRODUCTION

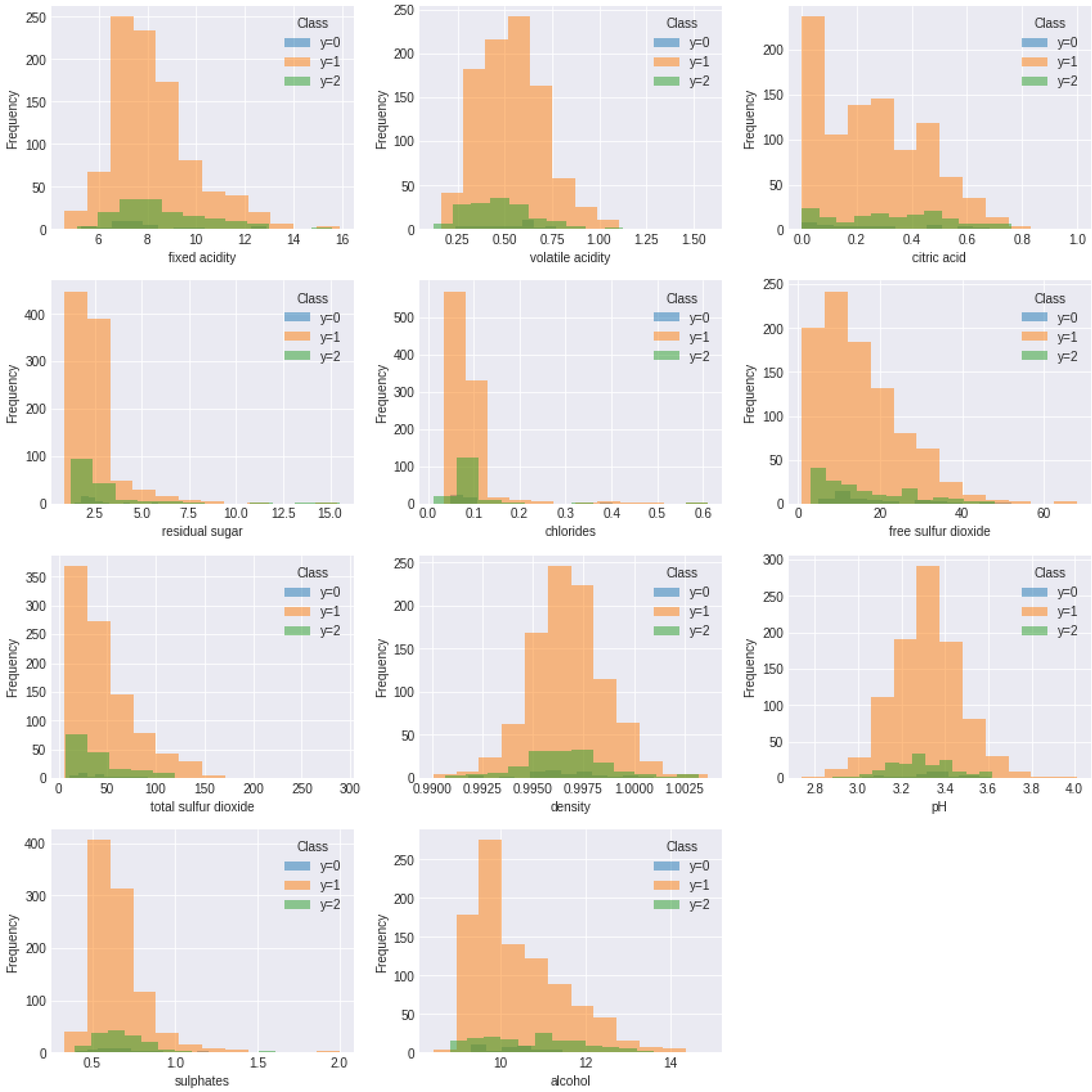
This study aims to assess the performance of two multi-class classification approaches on the "Wine Quality" dataset. With over 1000 samples of red wine and 11 attributes, the goal is to predict the wine quality using the provided "Y" variable. By evaluating the effectiveness of these methods, we can gain insights into their suitability for wine quality prediction, aiding in future model selection and application.

2. EXPLORATORY ANALYSIS

Histograms of all attributes, including the target variable "Y" (quality), provide a visual representation of data distribution and help identify patterns and outliers. Examining the histogram of "Y" allows assessment of class distribution, ensuring a balanced dataset for accurate classification. This analysis sets the foundation for further exploration and model development.



For a better understanding of the distribution of each variable for each different class, the following plot shows the distribution dividing by the Y value:

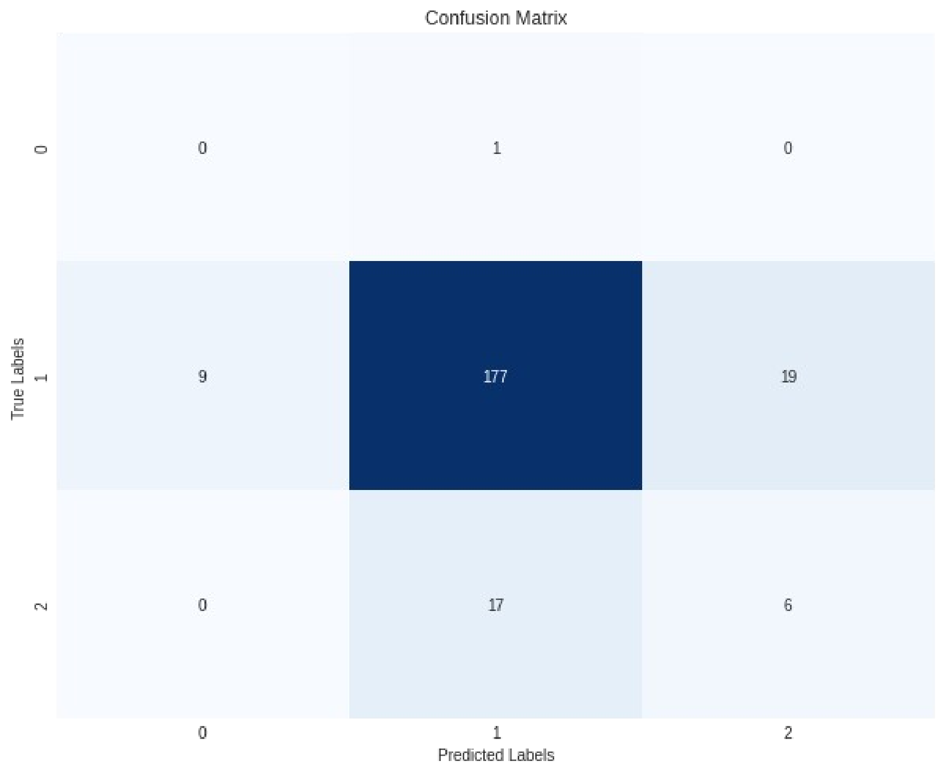


Once the proportion of the classes are not similar we can say that the classifier is unlikely to perform well. Furthermore, if the histograms of the attributes for the two classes are significantly different, it suggests that the attributes have different distributions for the two classes and are therefore informative for classification. On the other hand, if the histograms are similar, it suggests that the attributes are not strongly associated with the class labels and may not be useful for classification.

3. GAUSSIAN PROCESS CLASSIFIER

In this experiment, we employed the Gaussian Process Classifier with two multi-class classification approaches: One versus All (OvA) and One versus One (OvO). The goal was to calculate and compare the test accuracy, precision, and recall using the entire "Wine Quality" dataset. By utilizing the OvA and OvO classifiers, we were able to assess their performance in predicting wine quality across all classes. These results provide valuable insights into the effectiveness of each approach and their suitability for the wine quality prediction task. The results can be seen bellow:

	One Vs All	One Vs One
Accuracy	79.0%	80.0%
Recall	36.7%	38.2%
Precision	35.8%	37.5%



Based on the results obtained, both the One vs All (OvA) and One vs One (OvO) classifiers demonstrate similar levels of accuracy in predicting the wine quality on the "Wine Quality" dataset. The OvA classifier achieved an accuracy of 79% while the OvO classifier achieved 80%.

In terms of recall, which measures the ability to correctly identify positive samples, the OvO classifier outperformed the OvA classifier with a recall of 38.2% compared to 36.7%.

Similarly, in terms of precision, which measures the proportion of correctly predicted positive samples out of the total predicted positive samples, the OvO classifier exhibited better performance with a precision of 37.5% compared to 35.8% for the OvA classifier.

Based on these results, it can be concluded that the OvO classifier performs slightly better than the OvA classifier in terms of recall and precision, although the difference is relatively small. However, further analysis and evaluation may be required to fully understand the performance characteristics and potential trade-offs of each classifier in the context of the wine quality prediction task.