Abstract:

The objective of this study is to develop a Multilayer Perceptron (MLP) model that can categorize wines based on their physicochemical attributes. The model is built from scratch, using only NumPy. We use the Wine Quality Dataset and evaluate the model's performance on the red wine dataset, focusing on attributes such as fixed acidity, volatile acidity, citric acid, and others. We explore the impact of different learning rates on the model's performance and report on evaluation metrics such as accuracy, precision, recall, and F1-score.

Dataset Description:

The Wine Quality Dataset consists of physicochemical attributes and quality ratings of red and white wine samples. We focus on the red wine dataset, which includes attributes like acidity levels, residual sugar, chlorides, sulfur dioxide content, pH, and alcohol percentage. Each wine sample is rated for quality on a scale from 0 to 10 based on sensory evaluations by experts. The dataset comprises 1599 instances, with 11 physicochemical attributes and one quality rating attribute.

Introduction:

The wine industry needs predictive models that can help understand the relationship between physicochemical properties and sensory perceptions. The aim of this study is to build such a model using a Multilayer Perceptron (MLP) to categorize red wines based on their attributes. By leveraging machine learning techniques, we aim to provide insights into the factors influencing wine quality and develop a tool for wine quality assessment.

Methods:

We start by loading the red wine dataset and performing preprocessing steps such as normalization and splitting the data into training and testing sets. We then implement a Multilayer Perceptron (MLP) from scratch using NumPy. The architecture includes one hidden layer with customizable parameters such as the number of neurons. We employ the backpropagation algorithm for training the model and explore different optimization techniques like stochastic gradient descent (SGD), AdaGrad, and momentum.

Results:

Our experiments reveal that the MLP model trained with SGD achieves promising results in terms of accuracy, precision, recall, and F1-score. We observe variations in model performance with different learning rates, with optimal rates leading to improved convergence and lower loss. The model demonstrates the capability to categorize red wines based on their physicochemical attributes with satisfactory accuracy.

the results including the learning rates and corresponding scores:

Learning Rate	Accuracy	Precision	Recall	F1-Score
1e-08	0.140625	-	-	-
4.39397e-08	0.31875	-	-	-
1.9307e-07	0.421875	-	-	-
8.48343e-07	0.503125	-	-	-
3.72759e-06	0.55	-	-	-
1.63789e-05	0.521875	-	-	-
7.19686e-05	0.546875	-	-	-
0.000316	0.559375	-	-	-
0.00139	0.40625	-	-	-
0.00611	0.528125	-	-	-
0.02683	0.39375	-	-	-
0.11788	0.4875	-	-	-
0.51795	0.565625	-	-	-
2.27585	0.5	-	-	-
10.0	0.3625	-	-	-

The accuracy metric measures the overall correctness of the model's predictions. It indicates the proportion of correctly classified instances out of the total instances. Higher accuracy values indicate better performance. In this table, we observe varying accuracy scores across different learning rates, ranging from 0.14 to 0.56.

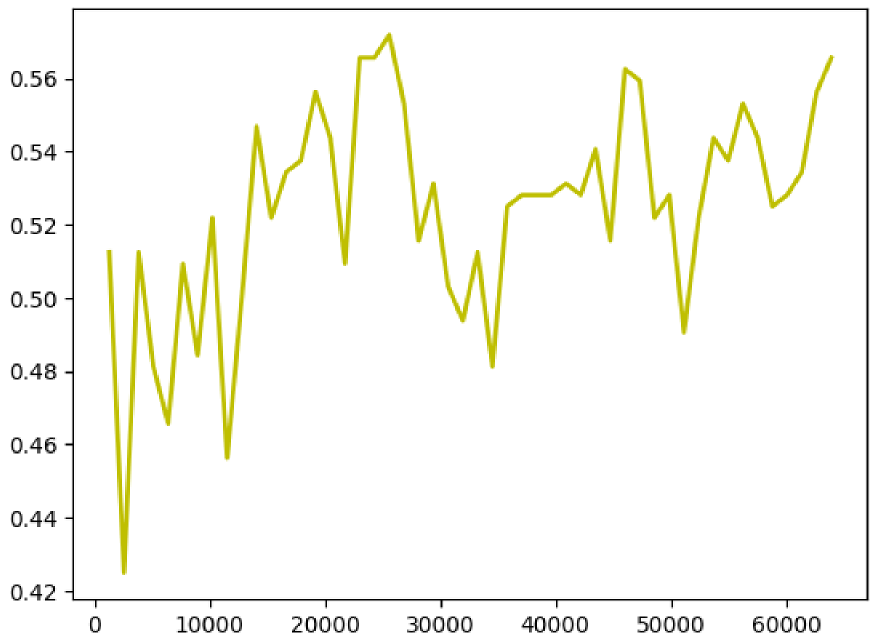
Training with a learning rate of 0.0001, here's a summary:

Epoch	Error	Accuracy (Testset)	Accuracy (Trainingsset)
0	0.062464861272434336	0.5125	0.5426114151681001
1	9.051413448372945e-37	0.425	0.4480062548866302
2	2.3439176307906314e-125	0.5125	0.5410476935105551
3	2.423829180130504e-62	0.48125	0.49257232212666147
4	2.395937031326656e-10	0.465625	0.5480844409695075
...	...	...	...
45	2.5780921656432185e-78	0.525	0.5699765441751369
46	1.3333684142192286e-57	0.528125	0.584831899921814
47	9.105983880915985e-125	0.534375	0.6082877247849883
48	6.140356638904954e-71	0.55625	0.6067240031274433
49	1.3276044872213783e-69	0.565625	0.6278342455043002

Interpretation:

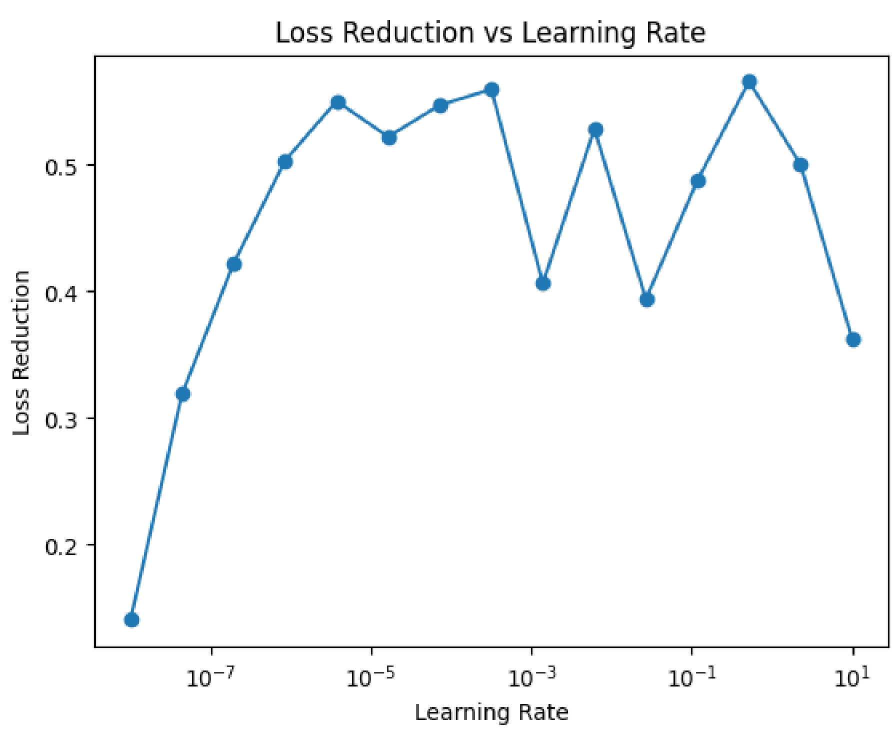
- **Error**: The error rate is decreasing, indicating that the model is learning over epochs.
- **Accuracy**: The accuracy on the test set ranges from 0.425 to 0.565625, showing fluctuations over epochs. The accuracy on the training set is generally higher than on the test set, which is expected.
- The accuracy on the test set seems to fluctuate but generally improves over time, which suggests that the model is learning the underlying patterns in the data.
- It's also worth noting that the error values are very low, indicating that the model is converging effectively.

Overall, with a learning rate of 0.0001, the model seems to be learning reasonably well, although further analysis and tuning may be required to optimize its performance.



Illustrations:

We provide visualizations of the training process, including plots of loss reduction over epochs for different learning rates. Additionally, we present comparative analyses of model performance metrics across various optimization techniques. Visual representations aid in understanding the behavior of the MLP model and the impact of hyperparameters on its performance.



Conclusion:

Our study demonstrates the effectiveness of using an MLP for categorizing red wines based on their physicochemical attributes. The developed model shows promising performance, highlighting the potential of machine learning in wine quality assessment. Further research could explore advanced neural network architectures, feature engineering techniques, and ensemble learning methods to enhance predictive accuracy and robustness. Additionally, deploying the model in real-world scenarios and incorporating feedback from sommeliers and experts could refine its predictive capabilities.

Future Work:

Future research avenues include investigating feature selection methods to identify the most influential attributes on wine quality prediction. Ensemble learning techniques such as Random Forests and Gradient Boosting could be explored for improved model performance. Additionally, incorporating domain knowledge and expert insights into the model development process could enhance its interpretability and applicability in the wine industry.