Task 1: Multinomial Regression

In this task, we implement a Multinomial logistic regression model from scratch using
and pandas. We train this model on Wine Quality Dataset to classify a wine's quality based on the values of its various contents.

Arnav Negi

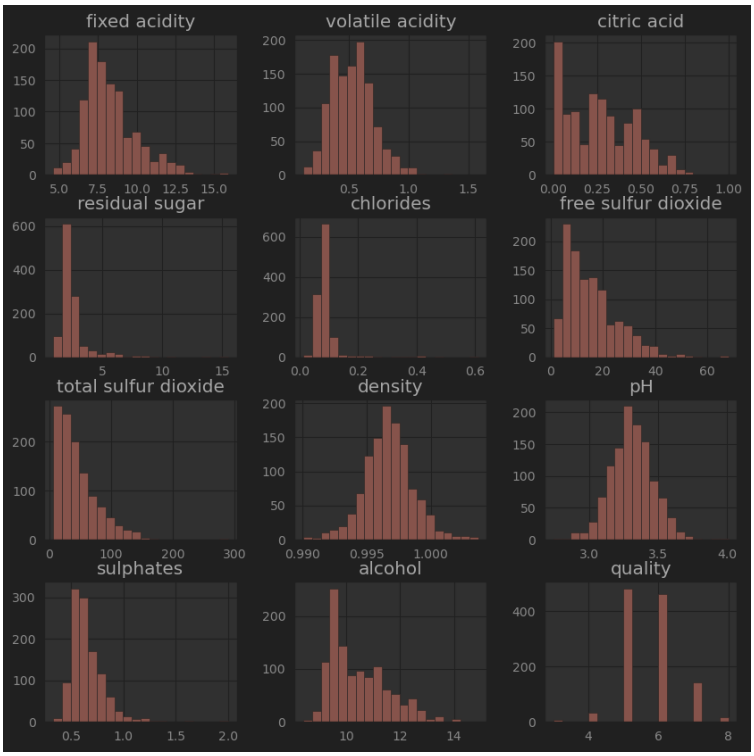
Data Exploration

The Wine Quality datset is as follows.

8 rows 8 rows × 12 columns pd.DataFrame CSV

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density
count	1143.000000	1143.000000	1143.000000	1143.000000	1143.000000	1143.000000	1143.000000	1143.00
mean	8.311111	0.531339	0.268364	2.532152	0.086933	15.615486	45.914698	0.99
std	1.747595	0.179633	0.196686	1.355917	0.047267	10.250486	32.782130	0.00
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	0.99
25%	7.100000	0.392500	0.090000	1.900000	0.070000	7.000000	21.000000	0.99
50%	7.900000	0.520000	0.250000	2.200000	0.079000	13.000000	37.000000	0.99
75%	9.100000	0.640000	0.420000	2.600000	0.090000	21.000000	61.000000	0.99
max	15.900000	1.580000	1.000000	15.500000	0.611000	68.000000	289.000000	1.00

The distribution of labels is given below,

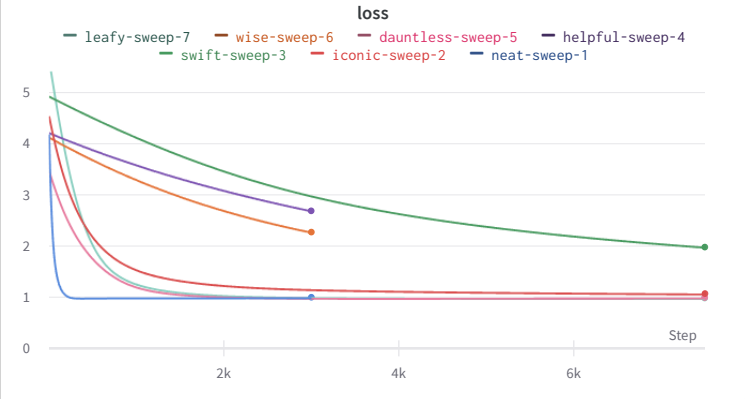


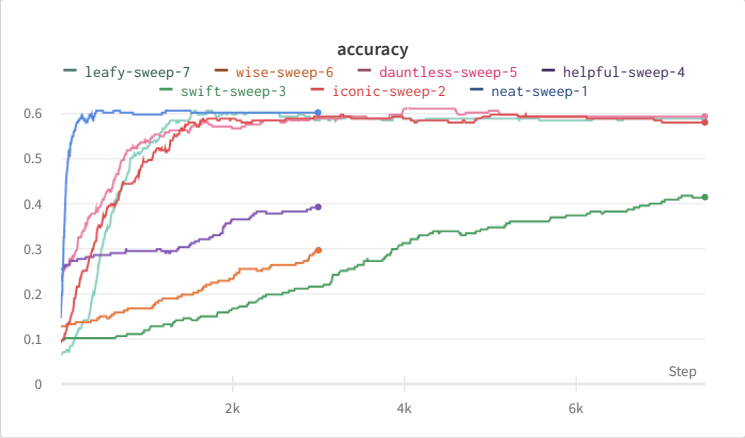
Model training and tuning

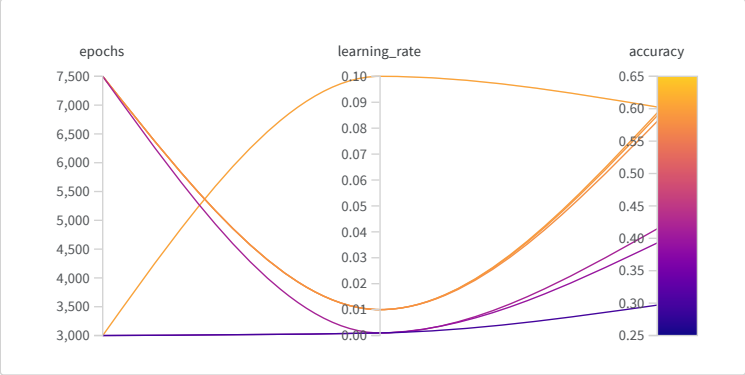
We train a multinomial regression model to classify the quality of the wine, given its data. We tune for the following:

- learning rate
- epochs

This gives the following results:







Run set 7

The best model gives 60.26% accuracy and has

- epochs = 3000
- learning rate = 0.1

https://wandb.ai/arnav-team/assignment-3/reports/Task-1-Multinomial-Regression--Vmlldzo1Nzg3MjAy