Logistic Regression Classifier Implementation on the Wine Dataset

Status Report 2

Tommy Liu and Pavel Filippov

November 23, 2025

Project Update

Last week, the project team finished the implementation of the LR Classifier and validated its correctness. This week, the team successfully integrated its LR Classifier into the Classifier Factory and ran some basic accuracy tests. The project team observed extraordinarily low accuracies with its classifier on multiclass classification via AVA/OVA. The team tried using different metrics for confidence and varying the number of iterations, but the accuracy remained low. For the next two days, the project team will validate the Logistic Regression part in the code of the AVA/OVA classifiers. Meanwhile, the team will research more information on whether logistic regression is suitable for text classification.

Results

Phase 2 of the project was to integrate the classifier into the classifier factory and perform basic experimentation. The main results the project team has are:

- When all 10 examples in the default dataset were used for training, the accuracy on the training set after a single iteration is 60%, meaning that 6/10 examples were classified correctly.

- The accuracy of the LR Classifier on *wine* dataset was around 50% using AVA and around 15% using OVA. The accuracies did not change much after tuning various parameters. The project team conjectures that the dataset might need feature scaling or normalization.

Problems

The project team has successfully integrated its LR Classifier into the Classifier Factory, and run some basic accuracy tests. The primary issue that has drawn the team's attention is that the LR Classifier has an unexpectedly low accuracy on text files. Even when the entire *wine* dataset was used for both training and testing, the accuracy remains around 50% using AVA and around 15% using OVA. There are two possibilities for the observed low accuracy: either the team did not correctly integrate its LR Classifier into AVA/OVA, or logistic regression is not a viable approach for text classification.

Hours

The team worked collaboratively for 4 hours on November 23, from 2:30PM - 5:30PM and from 7PM - 8PM.

Code

The repository is attached here.