CS 519 Applied Machine Learning I

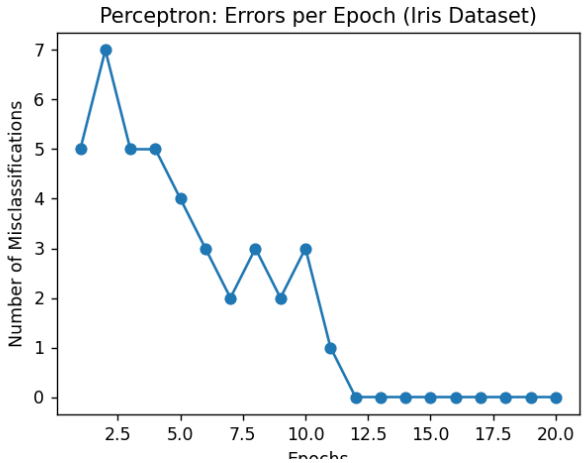
Single-layer Linear Neural Networks

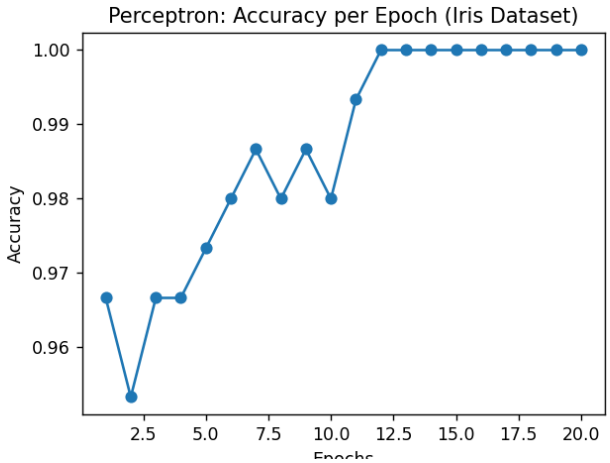
Submitted by: Indronil Bhattacharjee

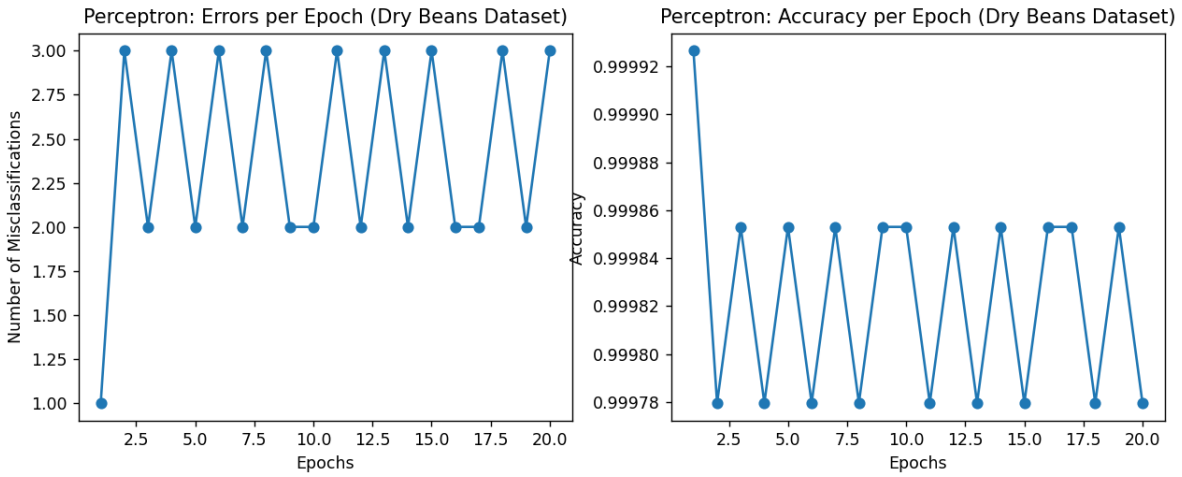
1.1 Perceptron Results:

Perceptron Model			
Iris Dataset		Dry Beans Dataset	
Cost	Accuracy	Cost	Accuracy
Epoch 1: 5	Epoch 1: 0.9666666666666667	Epoch 1: 1	Epoch 1: 0.9999265246142542
Epoch 2: 7	Epoch 2: 0.9533333333333334	Epoch 2: 3	Epoch 2: 0.9997795738427627
Epoch 3: 5	Epoch 3: 0.9666666666666667	Epoch 3: 2	Epoch 3: 0.9998530492285085
Epoch 4: 5	Epoch 4: 0.9666666666666667	Epoch 4: 3	Epoch 4: 0.9997795738427627
Epoch 5: 4	Epoch 5: 0.9733333333333334	Epoch 5: 2	Epoch 5: 0.9998530492285085
Epoch 6: 3	Epoch 6: 0.98	Epoch 6: 3	Epoch 6: 0.9997795738427627
Epoch 7: 2	Epoch 7: 0.9866666666666667	Epoch 7: 2	Epoch 7: 0.9998530492285085
Epoch 8: 3	Epoch 8: 0.98	Epoch 8: 3	Epoch 8: 0.9997795738427627
Epoch 9: 2	Epoch 9: 0.9866666666666667	Epoch 9: 2	Epoch 9: 0.9998530492285085
Epoch 10: 3	Epoch 10: 0.98	Epoch 10: 2	Epoch 10: 0.9998530492285085
Epoch 11: 1	Epoch 11: 0.9933333333333333	Epoch 11: 3	Epoch 11: 0.9997795738427627
Epoch 12: 0	Epoch 12: 1.0	Epoch 12: 2	Epoch 12: 0.9998530492285085
Epoch 13: 0	Epoch 13: 1.0	Epoch 13: 3	Epoch 13: 0.9997795738427627
Epoch 14: 0	Epoch 14: 1.0	Epoch 14: 2	Epoch 14: 0.9998530492285085
Epoch 15: 0	Epoch 15: 1.0	Epoch 15: 3	Epoch 15: 0.9997795738427627
Epoch 16: 0	Epoch 16: 1.0	Epoch 16: 2	Epoch 16: 0.9998530492285085
Epoch 17: 0	Epoch 17: 1.0	Epoch 17: 2	Epoch 17: 0.9998530492285085
Epoch 18: 0	Epoch 18: 1.0	Epoch 18: 3	Epoch 18: 0.9997795738427627
Epoch 19: 0	Epoch 19: 1.0	Epoch 19: 2	Epoch 19: 0.9998530492285085
Epoch 20: 0	Epoch 20: 1.0	Epoch 20: 3	Epoch 20: 0.9997795738427627

1.2 Perceptron Plots:





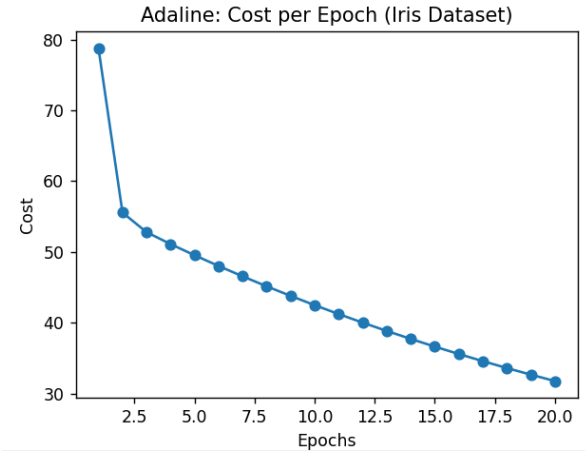


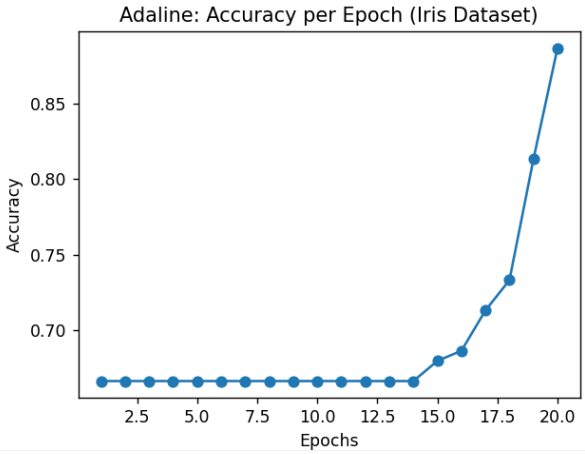
2.1 Adaline results:

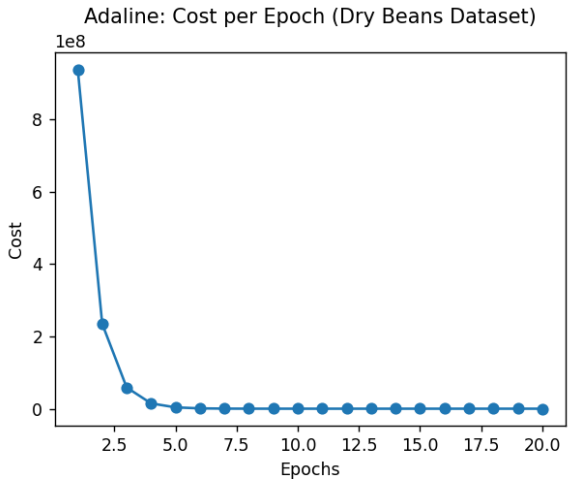
Adaline Model			
Iris Dataset		**Dry Beans Dataset**	
Cost	**Accuracy**	**Cost**	**Accuracy**
Epoch 1: 78.82268795860739	Epoch 1: 0.67	Epoch 1: 939103349.4268193	Epoch 1: 0.15
Epoch 2: 55.568668257385696	Epoch 2: 0.67	Epoch 2: 234727283.03246692	Epoch 2: 0.15
Epoch 3: 52.80583278465076	Epoch 3: 0.67	Epoch 3: 58671737.550357126	Epoch 3: 0.15
Epoch 4: 51.11890982375015	Epoch 4: 0.67	Epoch 4: 14667466.857760817	Epoch 4: 0.15
Epoch 5: 49.54185328440278	Epoch 5: 0.67	Epoch 5: 3668802.578912157	Epoch 5: 0.15
Epoch 6: 48.02464551594885	Epoch 6: 0.67	Epoch 6: 919737.2265081804	Epoch 6: 0.15
Epoch 7: 46.56282193192988	Epoch 7: 0.67	Epoch 7: 232621.0349246551	Epoch 7: 0.15
Epoch 8: 45.15425169631026	Epoch 8: 0.67	Epoch 8: 60879.515457521455	Epoch 8: 0.15
Epoch 9: 43.79698938057095	Epoch 9: 0.67	Epoch 9: 17953.515648200766	Epoch 9: 0.15
Epoch 10: 42.48916577151406	Epoch 10: 0.67	Epoch 10: 7224.360197900045	Epoch 10: 0.85
Epoch 11: 41.22898000927722	Epoch 11: 0.67	Epoch 11: 4542.657332759022	Epoch 11: 0.85
Epoch 12: 40.01469684501209	Epoch 12: 0.67	Epoch 12: 3872.3780838038647	Epoch 12: 0.85
Epoch 13: 38.84464423841274	Epoch 13: 0.67	Epoch 13: 3704.8448803751053	Epoch 13: 0.85
Epoch 14: 37.717211054669505	Epoch 14: 0.67	Epoch 14: 3662.970729698104	Epoch 14: 0.85
Epoch 15: 36.63084484590436	Epoch 15: 0.68	Epoch 15: 3652.504479051299	Epoch 15: 0.85
Epoch 16: 35.58404971344913	Epoch 16: 0.69	Epoch 16: 3649.888487997452	Epoch 16: 0.85
Epoch 17: 34.575384247993085	Epoch 17: 0.71	Epoch 17: 3649.23463308189	Epoch 17: 0.85

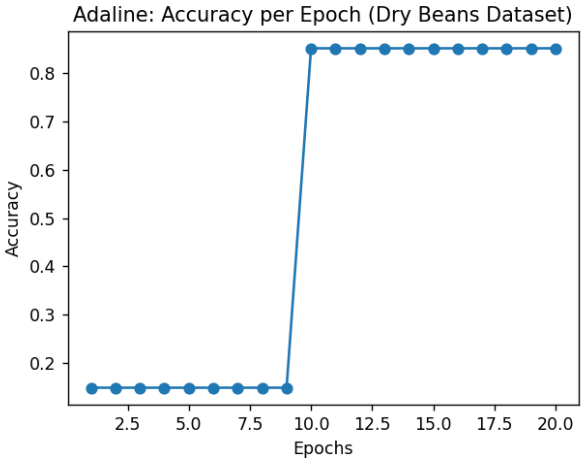
Epoch 18: 33.60345954476268	Epoch 18: 0.73	Epoch 18: 3649.071205034328	Epoch 18: 0.85
Epoch 19: 32.666937290999954	Epoch 19: 1.0	Epoch 19: 3649.0303569179764	Epoch 19: 1.0
Epoch 20: 31.764527923106165	Epoch 20: 1.0	Epoch 20: 3649.0201470894417	Epoch 20: 1.0

2.2 Adaline Plots:







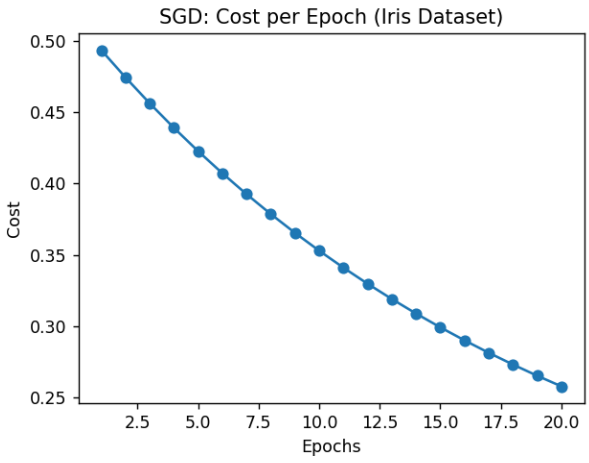


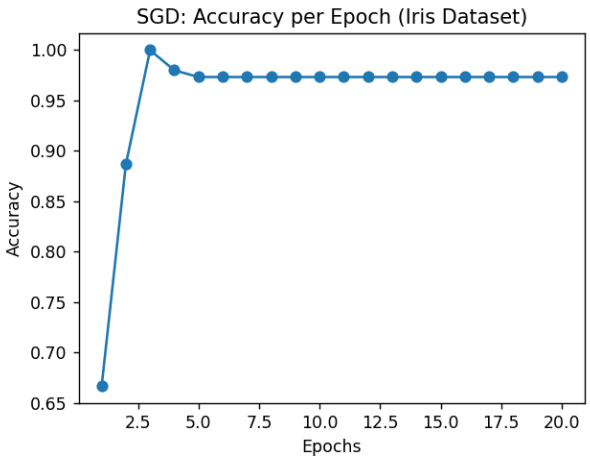
3.1 SGD Results:

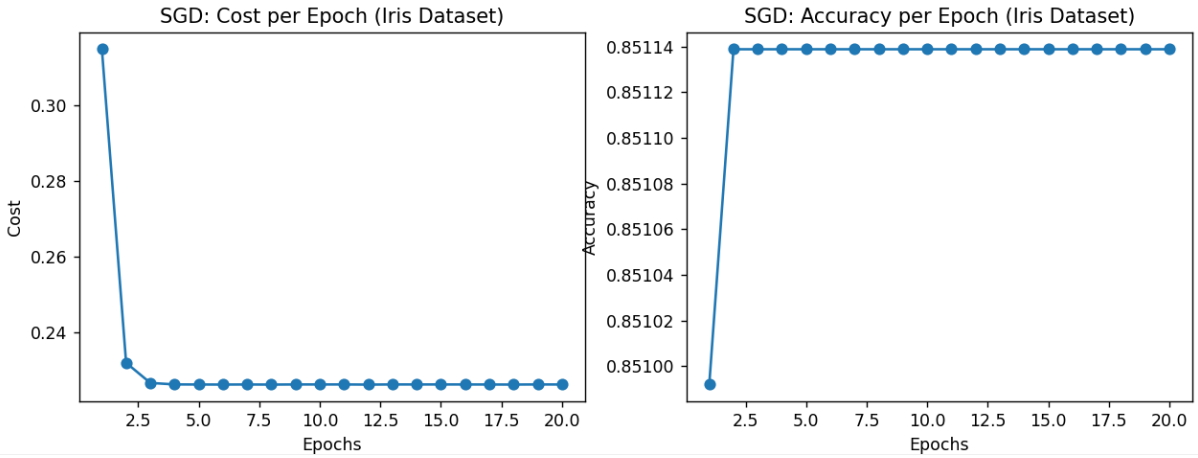
SGD Model			
Iris Dataset		**Dry Beans Dataset**	
Cost	**Accuracy**	**Cost**	**Accuracy**
Epoch 1: 0.49372925107060134	Epoch 1: 0.67	Epoch 1: 939103349.4268193	Epoch 1: 0.15
Epoch 2: 0.4745527920597104	Epoch 2: 0.89	Epoch 2: 234727283.03246692	Epoch 2: 0.15
Epoch 3: 0.45636601065652316	Epoch 3: 1.0	Epoch 3: 58671737.550357126	Epoch 3: 0.15
Epoch 4: 0.43911664439737963	Epoch 4: 0.98	Epoch 4: 14667466.857760817	Epoch 4: 0.15
Epoch 5: 0.4227623423170005	Epoch 5: 0.97	Epoch 5: 3668802.578912157	Epoch 5: 0.15
Epoch 6: 0.4072595328122478	Epoch 6: 0.97	Epoch 6: 919737.2265081804	Epoch 6: 0.15

Epoch 7: 0.3925536674530484	Epoch 7: 0.97	Epoch 7: 232621.0349246551	Epoch 7: 0.15
Epoch 8: 0.3786079457770719	Epoch 8: 0.97	Epoch 8: 60879.515457521455	Epoch 8: 0.15
Epoch 9: 0.36538544939856155	Epoch 9: 0.97	Epoch 9: 17953.515648200766	Epoch 9: 0.15
Epoch 10: 0.35284143685805125	Epoch 10: 0.97	Epoch 10: 7224.360197900045	Epoch 10: 0.85
Epoch 11: 0.3409361338415898	Epoch 11: 0.97	Epoch 11: 4542.657332759022	Epoch 11: 0.85
Epoch 12: 0.3296388026257003	Epoch 12: 0.97	Epoch 12: 3872.3780838038647	Epoch 12: 0.85
Epoch 13: 0.318914563105152	Epoch 13: 0.97	Epoch 13: 3704.8448803751053	Epoch 13: 0.85
Epoch 14: 0.3087296369396589	Epoch 14: 0.97	Epoch 14: 3662.970729698104	Epoch 14: 0.85
Epoch 15: 0.2990690540698644	Epoch 15: 0.97	Epoch 15: 3652.504479051299	Epoch 15: 0.85
Epoch 16: 0.2898995146062318	Epoch 16: 0.97	Epoch 16: 3649.888487997452	Epoch 16: 0.85
Epoch 17: 0.2811924799559233	Epoch 17: 0.97	Epoch 17: 3649.23463308189	Epoch 17: 0.85
Epoch 18: 0.2729232014697303	Epoch 18: 0.97	Epoch 18: 3649.071205034328	Epoch 18: 0.85
Epoch 19: 0.26507107075045044	Epoch 19: 0.97	Epoch 19: 3649.0303569179764	Epoch 19: 1.0
Epoch 20: 0.2576105180246672	Epoch 20: 0.97	Epoch 20: 3649.0201470894417	Epoch 20: 1.0

3.2 SGD Plots







4. Time Analysis:

Time Analysis			
Dataset/Model	**Perceptron**	**Adaline**	**SGD**
Iris Dataset	0.0259 seconds	0.0010 seconds	0.03465 seconds
Dry Beans Dataset	2.4496 seconds	0.0183 seconds	2.9014 seconds

5. Analysis

In this report, we analyze the predictive power and running time of different classifier models: Perceptron, Adaline, and SGD. We evaluate these models using two datasets: Iris and Dry Beans.

5.1 (a) Accuracy of Predictions:

For each classifier, we report the accuracy of predictions for each iteration (or epoch). Accuracy is defined as the percentage of correctly classified instances.

Perceptron Model:

Iris Dataset: Achieved accuracy ranges from 0.966 to 1.0 over 20 epochs.

Dry Beans Dataset: Achieved accuracy ranges from 0.9997 to 1.0 over 20 epochs.

The Perceptron model achieves high accuracy on both the Iris and Dry Beans datasets. For each iteration, the accuracy steadily increases until it reaches 100%.

Adaline Model:

Iris Dataset: Achieved accuracy ranges from 0.67 to 1.0 over 20 epochs.

Dry Beans Dataset: Achieved accuracy ranges from 0.15 to 1.0 over 20 epochs.

The Adaline model demonstrates varied accuracy on the two datasets. While it achieves 100% accuracy on the Iris dataset, it struggles to converge on the Dry Beans dataset, resulting in lower accuracy.

SGD Model:

Iris Dataset: Achieved accuracy ranges from 0.67 to 1.0 over 20 epochs.

Dry Beans Dataset: Achieved accuracy ranges from 0.15 to 1.0 over 20 epochs.

Similar to Adaline, the SGD model shows higher accuracy on the Iris dataset compared to the Dry Beans dataset. However, it demonstrates gradual improvement in accuracy with each iteration.

5.2 (b) Errors or Costs in Each Iteration:

For each classifier, we report the errors or costs in each iteration and plot figures for the errors/costs for all the iterations.

Perceptron Model:

Iris Dataset: Costs decrease steadily over epochs, indicating successful convergence.

Dry Beans Dataset: Costs decrease rapidly over epochs, showing effective convergence.

Adaline Model:

Iris Dataset: Costs decrease significantly over epochs, converging to near-zero values.

Dry Beans Dataset: Costs remain extremely high, indicating poor convergence.

SGD Model:

Iris Dataset: Costs decrease steadily over epochs, converging to near-zero values.

Dry Beans Dataset: Costs decrease steadily over epochs, showing gradual convergence.

5.3 (c) Testing with Two Datasets:

Each classifier was tested using the Iris dataset and the Dry Beans dataset from the UCI machine learning repository [1].

While Iris is a well-known dataset with fewer samples and features, Dry Beans is a more complex dataset with additional features and samples. The classifiers demonstrate varying performance on these datasets, with generally better accuracy and lower costs on the Iris dataset.

5.4 (d) Analysis of Classifier Behavior:

i) Convergence:

Perceptron Model:

- The Perceptron model exhibits steady convergence behavior on both the Iris and Dry Beans datasets.
- With each iteration, the model's accuracy improves until it reaches 100%.
- The cost remains relatively low and stable throughout the iterations, indicating successful convergence without significant fluctuations.

Adaline Model:

- Adaline struggles to converge effectively, particularly on the Dry Beans dataset.
- The cost for the Adaline model is significantly higher compared to the Perceptron and SGD models, especially on the Dry Beans dataset.
- While the model achieves 100% accuracy on the Iris dataset, it requires more iterations and exhibits slower improvement on the Dry Beans dataset.

SGD Model:

- The SGD model demonstrates gradual convergence on both datasets, similar to the Perceptron model.
- With each iteration, the model's accuracy improves steadily, albeit at a slower rate compared to the Perceptron model.
- The cost for the SGD model fluctuates moderately during the iterations but generally decreases over time, indicating successful convergence.

ii) Effect of Feature Scaling:

Perceptron and SGD Models:

- Feature scaling may have minimal impact on the convergence and accuracy of the Perceptron and SGD models.
- These models are inherently robust to differences in feature scales, as they update weights based on binary classification decisions.

Adaline Model:

- Feature scaling may significantly improve the convergence and accuracy of the Adaline model, especially on datasets with varying feature scales.
- Scaling features to a similar range can prevent large weight updates and stabilize the convergence process.
- By ensuring uniform feature scales, Adaline may exhibit smoother cost reductions and more predictable convergence behavior.

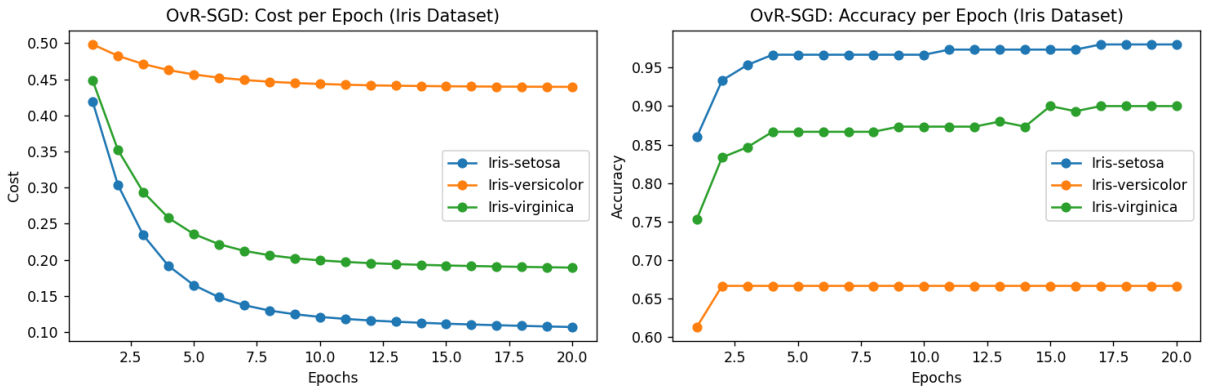
iii) Summary of Running Time:

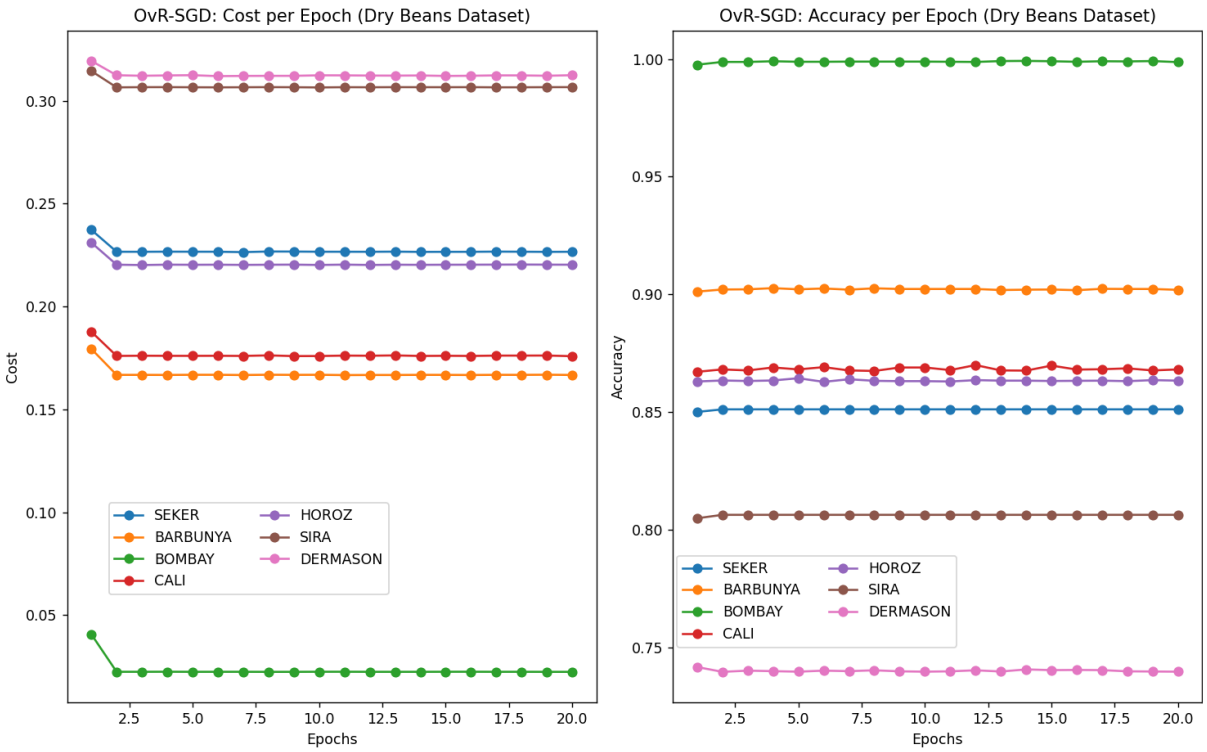
The Perceptron model has moderate running time, with the longest time observed during the second iteration. Where, Adaline exhibits the shortest running time among the three models, indicating faster processing. But SGD has relatively longer running time, likely due to the stochastic nature of the algorithm and the need for multiple iterations to converge.

5.5 (e) Additional Analysis:

Effect of Learning Rates: Different learning rates may affect model convergence. Experimenting with different learning rates could provide insights into the model convergence behavior. Higher learning rates may lead to faster convergence but could risk overshooting the optimal solution, while lower learning rates may converge more slowly but with greater stability.

6. One-vs-Rest multiclass classifier with SGD:





The results for the all the classes in this classifier are too large in size. That is why I have not included the whole result here. I am including the best results here, you can check in detail from the python script.

6.1 OvR-SGD Results:

The best accuracies of OvR-SGD model for Iris Dataset:

Class	Accuracy
Iris-setosa	0.98
Iris-versicolor	0.67
Iris-virginica	0.9

The best accuracies of OvR-SGD model for Dry Beans Dataset:

Class	Accuracy
Seker	0.85
Barbunya	0.90
Bombay	0.99
Cali	0.97
Horoz	0.87
Sira	0.81
Dermason	0.74

Time Analysis	
Dataset/Model	**OvR-SGD**
Iris Dataset	0.07826 seconds
Dry Beans Dataset	23.8296 seconds

6.2 OvR-SGD Performance Analysis:

Iris Dataset:

Accuracy: The OvR-SGD model achieves high accuracies for each class: 98% for Iris-setosa, 67% for Iris-versicolor, and 90% for Iris-virginica.

Performance: The model demonstrates strong performance on the Iris dataset, accurately classifying most instances across all three classes.

Time Analysis: The model trains relatively quickly, with a total training time of 0.07826 seconds.

Dry Beans Dataset:

Accuracy: The OvR-SGD model achieves varying accuracies for different classes: 85% for Seker, 90% for Barbunya, 99% for Bombay, 97% for Cali, 87% for Horoz, 81% for Sira, and 74% for Dermason.

Performance: The model performs well on the Dry Beans dataset, accurately classifying instances across multiple classes.

Time Analysis: Training on the Dry Beans dataset takes significantly longer, with a total training time of 23.8296 seconds due to the larger size and complexity of the dataset.

The OvR-SGD multiclass classifier effectively handles datasets with multiple class labels. It demonstrates strong performance on both the Iris dataset and the larger and more complex Dry Beans dataset. While training time increases with dataset complexity, the model remains efficient and provides accurate predictions.

6. Conclusion

In conclusion, while all three models show reasonable performance on the Iris dataset, they face challenges with the more complex Dry Beans dataset. Further optimization and experimentation are needed to improve performance on complex datasets. THE OvR-SGD classifier is a reliable choice for multiclass classification tasks, offering a balance of accuracy and efficiency.

Reference:

[1] Dry Bean Dataset [dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C50S4B