CSST102 – Machine Problem No. 2
Evaluating Machine Learning Model Performance Using Logistic Regression

---

1. Confusion Matrix Interpretation

The confusion matrix (see *confusion_matrix.png*) illustrates how well the model classified each Iris species in the test set. Each row represents the true label, and each column represents the predicted label. A perfect model would have all values along the diagonal and zeros elsewhere. In our results, most predictions fall along the diagonal, meaning the model correctly predicts the species in most cases. This indicates that Logistic Regression effectively separates the classes in the feature space.

Performance Metrics:

- Accuracy: 0.9333
- Precision (macro avg): 0.9333
- Recall (macro avg): 0.9333
- F1-Score (macro avg): 0.9333

High precision and recall show that the model not only performs well overall but also treats each class fairly. This fairness is crucial in multi-class classification tasks.

---

2. Cross-Validation Consistency

The model was evaluated using **5-Fold Cross Validation**, which divides the dataset into five parts, training and evaluating the model five times. The mean cross-validation accuracy was **0.9600**, with a standard deviation of **0.0435**. The low standard deviation indicates that the model's performance is consistent across different folds, suggesting that it is stable and not overly dependent on specific training samples.

This consistency demonstrates that the model's success is not due to chance on a single data split but rather a reflection of its general reliability across multiple partitions.

---

3. Learning Curve Insights

The learning curve (see *learning_curve.png*) shows the relationship between accuracy and the number of training samples used. It includes two lines:

- **Training Score:** Model accuracy on the training data.
- **Cross-Validation Score:** Model accuracy on unseen validation folds.

When the training dataset is small, the model achieves high training accuracy but struggles to generalize, resulting in a lower validation score. As the training size increases, both scores become closer, indicating that the model improves its generalization. When the two curves converge and plateau, it suggests that the model is balanced, neither overfitting nor underfitting.

Overall, the Logistic Regression model for the Iris dataset demonstrates good generalization and appropriate fitting behavior.

---

4. Possible Improvements

Although the model already performs well, several enhancements can be explored:

- **Feature Engineering:** Add polynomial feature interactions (e.g., `petal_length × petal_width`) to capture nonlinear relationships.
- **Hyperparameter Tuning:** Adjust the regularization strength (`C`) in Logistic Regression to balance bias and variance.
- **Alternative Models:** Compare results with other algorithms such as Decision Tree, K-Nearest Neighbors (KNN), or Support Vector Machine (SVM), as suggested in the challenge section.

Conclusion

By following the required workflow including data preprocessing, 80/20 data split, Logistic Regression training, 5-fold cross-validation, confusion matrix evaluation, and learning curve analysis, it is evident that Logistic Regression performs effectively and consistently on the Iris dataset. The model achieves high accuracy and stability, confirming its suitability for this type of classification problem.