

Detailed Summary

Iris Dataset

- Training Accuracy: 0.975 (97.5%)
- Test Accuracy (explicit parameters): 0.967 (96.7%)

Interpretation: The KNN classifier performed extremely well on the iris dataset. The training accuracy is close to perfect, and the test accuracy shows that the model generalizes very well to unseen data. The scatter plot of sepal length vs. sepal width confirms clear separation among species.

Simulated Dataset

- Training Accuracy: 0.942 (94.2%)
- Test Accuracy (explicit parameters): 0.900 (90.0%)

Interpretation: The synthetic dataset was more challenging because class boundaries overlapped. Still, KNN achieved strong performance. The training accuracy being higher than the test accuracy suggests maybe a small degree of overfitting, but overall the model maintained robust predictive ability. The scatter plot highlights the overlap in class clusters, explaining the lower accuracy compared to the iris dataset.

Decision Boundary (Simulated Dataset)

- The decision boundary plot demonstrates how KNN partitions the feature space into regions associated with each class.
- Regions are irregular and adapt to the distribution of training points, showing the non-parametric flexibility of KNN.
- This visualization helps explain why there are quite a few misclassifications near boundary areas where classes have overlapped.

Key Insights

- KNN is highly effective on clean, well-separated datasets like iris, reaching near-perfect accuracy.
- On more complex datasets with overlapping features, KNN still performs well but with slightly reduced accuracy.
- Explicit parameter tuning (e.g., `n_neighbors=5`, Euclidean distance) does not drastically change results but ensures stability.