

Machine Learning Assignment 3

1. Problem Definition

This project addresses a multi-class classification problem using the **Iris Dataset**, obtained from the "sklearn.datasets" module. The objective is to classify flowers into three species:

- Setosa
- Versicolor
- Virginica

We evaluate and compare the classification performance using two machine learning models:

- Support Vector Machine (SVM)
- Neural Network (NN)

2. Data Preprocessing

Dataset Summary:

- **Features:** 4 numerical input features (sepal length, sepal width, petal length, petal width)
- **Target Classes:** 3 classes (Setosa, Versicolor, Virginica)
- **Total Records:** 150

Preprocessing Steps:

- **Missing Values:** Checked and confirmed none.
- **Train-Test Split:** 70% training, 30% testing
- **Normalization:** Standardization applied using StandardScaler
- **Exploratory Data Analysis:**
 - Correlation heatmap was plotted.
 - Additional EDA (histograms, pair plots) recommended for future improvements.

3. SVM Implementation

Models Tested: Linear, Polynomial, RBF

Performance (Untuned Models): Accuracy ranged from ~81% to ~95% depending on the models

Hyperparameter Tuning: Using "GridSearchCV" on SVM with

- ✓ models: linear, poly, rbf
- ✓ C values: 0.1, 1, 10, 100
- ✓ Gamma values: 1, 0.1, 0.01, 0.001

Best Configuration:

- ✓ **Model:** Linear
- ✓ **C:** 10
- ✓ **Gamma:** 0.01
- ✓ **Accuracy:** 98.25%

Evaluation Metrics (Linear Model):

- **Confusion Matrix:** High true positives and true negatives.
- **Precision, Recall, F1-Score:** Excellent across classes.

4. Neural Network Implementation

Model Architecture:

- Input Layer: 4 neurons
- Hidden Layer: 1 hidden layer (16 neurons, ReLU activation or SIGMOID activation)
- Output Layer: 3 neurons (Softmax activation)

Training Details:

- Optimizer: Adam
- Loss Function: Categorical Crossentropy
- Epochs: 100

Performance:

- **Accuracy:** 86%
- **Loss Curve:** Smooth convergence
- **Confusion Matrix:** Strong results, comparable to SVM
- **Classification Report:** Balanced precision and recall

5. Model Comparison and Analysis

ROC Curves: Both models showed high AUC values, indicating strong discriminatory power.

Overfitting/Underfitting:

- No significant overfitting observed
- Smooth loss and performance curves

Improvement Suggestions:

- Try deeper neural networks or dropout layers
- Use more extensive EDA
- Apply feature selection for optimization

6. Conclusion

Both models performed excellently, with the **SVM slightly outperforming** the Neural Network in accuracy and recall. The results confirm that both are suitable classifiers for this multi-class flower classification problem. Future work can focus on model robustness and interpretability.

Team Members

ID	Name	Role
23010036	Farah Walid	Neural Network & Report
23011052	Basmala Hossam El-Din	Preprocessing, SVM & Markdowns
23010014	Rodina Mohamed	Model comparison & Analysis