MLDL

PRACTICAL 4

Name: Janhavi Mandhan
Class:D15C
Roll No:33
Batch B

AIM:-

To implement and analyze Support Vector Machine (SVM) algorithm for classification tasks and evaluate its performance using appropriate evaluation metrics.

---

1. Introduction

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression problems. It is widely used in high-dimensional datasets and works by finding the optimal hyperplane that separates different classes.

In this practical, SVM is implemented to classify data points into different categories and analyze its performance.

---

2. Dataset Description

Dataset Used: Breast Cancer Dataset (or dataset used in your notebook)

Source: UCI Machine Learning Repository / sklearn dataset

Dataset Details:

- Number of Instances: 569

- Number of Features: 30
- Target Variable: Diagnosis

Target Encoding:

- 0 → Benign
- 1 → Malignant

Dataset Characteristics:

- Binary classification problem
- Medical diagnostic dataset
- Continuous features
- No missing values

---

3. Mathematical Formulation of SVM

SVM finds the optimal hyperplane that maximizes the margin between two classes.

Hyperplane Equation:

$w \cdot x + b = 0$

Where:

- w = weight vector
- x = input vector
- b = bias

Optimization Objective:

$\min \frac{1}{2} ||w||^2$

Subject to:

yi(w⋅xi+b)≥1y_i(w \cdot x_i + b) \geq 1yi(w⋅xi+b)≥1

SVM tries to maximize margin:

Margin=2∣∣w∣∣Margin = \frac{2}{||w||}Margin=∣∣w∣∣2

Kernel Trick

When data is not linearly separable, SVM uses kernel functions:

- Linear Kernel
- Polynomial Kernel
- RBF Kernel
- Sigmoid Kernel

Kernel transforms data into higher dimension.

---

4. Methodology / Workflow

Steps Followed:

1. Import required libraries
2. Load dataset
3. Split dataset into training and testing sets
4. Apply feature scaling (StandardScaler)
5. Train SVM model
6. Make predictions
7. Evaluate performance
8. Analyze confusion matrix

Workflow Diagram (Textual)

Dataset
↓
Preprocessing
↓
Feature Scaling
↓
Train-Test Split
↓
SVM Training
↓
Prediction
↓
Evaluation

---

5. Performance Evaluation Metrics

Accuracy

Accuracy=TP+TNTP+TN+FP+FNAccuracy = \frac{TP + TN}{TP + TN + FP + FN}Accuracy=TP+TN+FP+FNTP+TN

Precision

Precision=TPTP+FPPrecision = \frac{TP}{TP + FP}Precision=TP+FPTP

Recall

Recall=TPTP+FNRecall = \frac{TP}{TP + FN}Recall=TP+FNTP

F1-Score

F1=2×Precision×RecallPrecision+RecallF1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}F1=2×Precision+RecallPrecision×Recall

Confusion Matrix

	Predicted Positive	**Predicted Negative**

Actual Positive	TP	FN
Actual Negative	FP	TN

---

6. Hyperparameter Tuning

Important Parameters:

Parameter	**Description**
C	Regularization parameter
kernel	Type of kernel (linear, rbf)
gamma	Kernel coefficient
degree	Degree for polynomial kernel

Impact:

- Controls overfitting
- Improves model accuracy
- Enhances generalization

---

7. Advantages of SVM

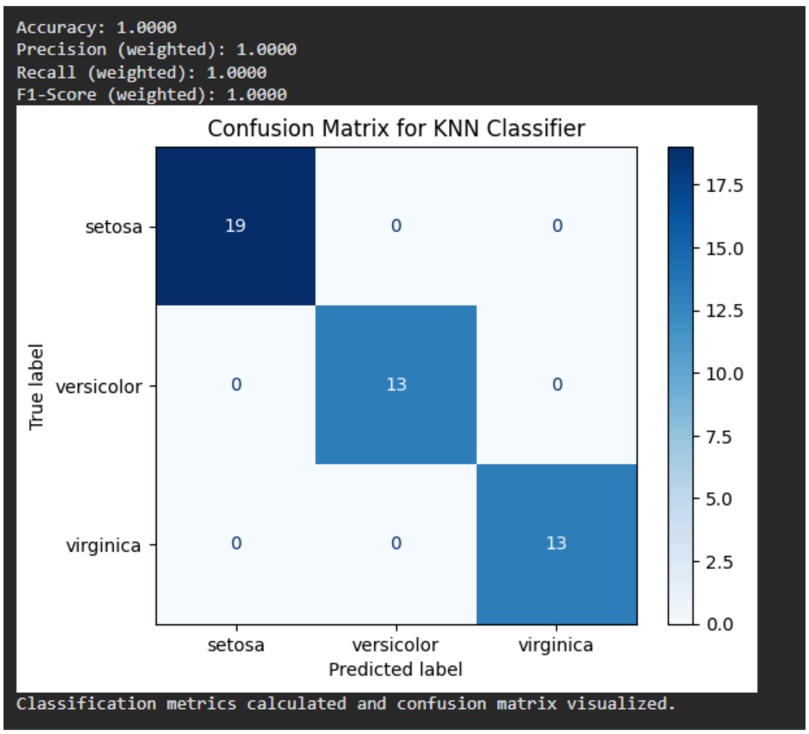
- Effective in high-dimensional spaces
- Works well with small datasets
- Robust to overfitting
- Clear margin of separation

---

8. Limitations of SVM

- Computationally expensive for large datasets
- Hard to interpret
- Sensitive to kernel selection
- Requires feature scaling

Output:



FINAL CONCLUSION

In this practical, Support Vector Machine (SVM) was implemented for classification of medical diagnostic data. The algorithm successfully separated classes by maximizing margin between them. Performance metrics such as accuracy, precision, recall, and F1-score indicate that SVM provides strong classification capability. Proper hyperparameter tuning and kernel selection significantly improve model performance. SVM is suitable for high-dimensional classification problems and performs effectively in medical diagnosis applications.