

22/NOV/23

AIM Assignment - 5

SVM classifier

Aim: Write a program to implement SVM classifier, and compare it with a decision tree algorithm.

Purpose: To study the SVM classifier and decision tree algorithms.

Alg: → SVM algorithm
→ Decision Tree Algorithm

platform: Windows / Arch-Linux (64 bit)
vs code, Python 3.12

Theory :

(I) Overview of SVM

→ SVM is "Supervised Machine Learning" algorithm that is used for machine learning and classification and regression.

→ It works by finding the hyperplane that best separates data into separate classes.

→ Kernel functions enable SVM to handle non-linear relationships.
e.g., Linear, Radial Basis Function.

(★)

Decision Tree

- They are tree like models which each node represent a decision based on input features.
- They recursively split data to form a tree structure, making decisions at each node.
- They are suitable for both classification and regression tasks.

→

Applications

(1)

Business Decision Support

e.g. Determining the optimal marketing strategy based on customer attributes.

(2)

Medical diagnosis

e.g. Predicting if a patient has certain disease based on symptoms.

(★)

Confusion Matrix

- It is a table that describes the performance of a classification algorithm.
- It shows True positive, True negative, False positive and False Negative values.

eg :

$$\begin{bmatrix} 23 & 7 \\ 10 & 60 \end{bmatrix}$$

for good Model

$$\begin{bmatrix} 23 & 7 \\ 58 & 12 \end{bmatrix}$$

for bad Model



Evaluation Metrics:

- (1) Accuracy: Overall correctness of the classifier
- (2) F1-score: Harmonic Mean of Precision and Recall, suitable for imbalanced datasets.
- (3) Recall (Sensitivity): Proportion of actual positives correctly predicted.



IRIS Dataset

- It is a widely used dataset in machine learning, containing measurements of IRIS flowers.
- There are 3 classes:
Setosa, Versicolor, Virginica

* FAQ :

Q(1) State the objectives of Lab Experiment:

→ To understand SVM and decision tree classifiers.

→ Evaluate and compare their performance using appropriate metrics

Q(2) Specify the parameters and hyperparameters used for both the SVM and decision tree classifiers (e.g. kernel type and SVM, max depth for decision tree)

→ SVM.

1. Kernel Type: Determines the type of decision boundary. Common options include:

(A) Linear: suitable for linearly separable data

(B) Radial Basis Function: (RBF)

Handles non linear relationships effectively.

2. Regularization Parameter (C): Controls the trade-off between achieving a smooth decision boundary and classifying training points correctly.

→ Decision Tree

① Maximum Depth: Controls the number of nodes and splits.

② Minimum samples split: The minimum no. of samples required to split the internal node.

③ Criterion: Function to measure quality of a split. Common options include:

A. (Gini Index) for impurity

B. (Entropy) for information gain

Q.3 Explain the importance of comparing these two algorithms:

→ ① Algorithm selection for specific task:

→ understanding how both algorithms work guides us to select the right alg on different types of data

② Performance Evaluation:

→ Comparison provides insights into strengths and weaknesses of each algorithm.

③ Resource Efficiency: Comparing computational efficiency, memory usage, etc helps in selecting the right algorithm.

④ Interpretability and Explainability:

→ Decision trees allow easy understanding of the process, and provide a transparent model.

→ SVM is less interpretable, but is more powerful

⑤ Handling Non-linearity:

⑥ Robustness to Outliers and Noise:

~~July 10
2023~~