Support Vector Machine (SVM) - Full Explanation

1. What is SVM?

Support Vector Machine (SVM) is a supervised machine learning algorithm mainly used for classification, but it can also be used for regression. It works by finding the best decision boundary (hyperplane) that separates different classes in the dataset.

2. Key Terms in SVM

(a) Hyperplane

The hyperplane is the decision boundary that separates data points of different classes. In 2D it's a line, in 3D it's a plane, and in higher dimensions it's called a hyperplane.

(b) Support Vectors

Support vectors are the data points that lie closest to the decision boundary. They are critical because the hyperplane is defined using them.

(c) Margin

The margin is the distance between the hyperplane and the nearest data points (support vectors). SVM aims to maximize this margin.

(d) Kernel Trick

When data is not linearly separable, SVM uses kernel functions to project it into higher dimensions where it can be separated. Common kernels are: Linear, Polynomial, RBF (Radial Basis Function), and Sigmoid.

(e) Soft Margin vs Hard Margin

Hard Margin: No misclassification allowed, works only with perfectly separable data. Soft Margin: Allows some misclassification for better generalization. Controlled by parameter C.

(f) Parameters

C (Regularization parameter): Controls trade-off between maximizing margin and minimizing classification error. High C = stricter, low C = more tolerance. Gamma (for RBF kernel): Defines influence of training points. High gamma = tighter boundary, low gamma = smoother boundary.

3. How SVM Works – Step by Step

1. Plot data points in feature space. 2. Choose a hyperplane that separates classes. 3. Identify support vectors. 4. Maximize the margin. 5. If not linearly separable, apply kernel trick. 6. Tune parameters C

and gamma for performance.

4. Advantages of SVM

- Works well with high-dimensional data. - Effective with non-linear boundaries using kernels. - Resistant to overfitting with proper parameters.

5. Disadvantages of SVM

- Slow for very large datasets. - Requires careful choice of kernel and parameters. - Doesn’t perform well with noisy or heavily overlapping data.

Summary

SVM finds the optimal hyperplane using support vectors to maximize the margin between classes. With the kernel trick, it can handle both linear and non-linear data effectively.