Introduction

- Support Vector Machines (SVMs) are a type of supervised learning algorithm used for classification and regression tasks.
- The main idea behind SVMs is to find a decision boundary (hyperplane) that not only separates the classes but also maximizes the distance between the closest training points and the hyperplane.
- This hyperplane is called the maximum-margin hyperplane, and the distance between the hyperplane and the closest training points is called the margin.
- By maximizing the margin, SVMs aim to create a robust classifier that can generalize well to unseen data.

Benefits of SVMs

1. **Robustness to Outliers**: SVMs are relatively insensitive to outliers due to the margin maximization process.
2. **Handling Non-linear Data**: SVMs can effectively handle non-linear data by using a technique called the kernel trick, which maps the data into a higher-dimensional space where it becomes linearly separable.
3. **Applicability to Classification and Regression**: SVMs can be used for both classification and regression tasks.

Mathematics and Logic

1. **Equation of the Hyperplane**:
    - $W_1X_1 + W_2X_2 + \ldots + W_nX_n + W_0 = 0$
    - $\mathbf{W} \cdot \mathbf{X} + W_0 = 0$
    - $\mathbf{W}^T\mathbf{X} + W_0 = 0$, where $\mathbf{W}^T$ is the transpose of the vector $\mathbf{W}$.
    - $\mathbf{W}^T\mathbf{X} = 0$, assuming that the hyperplane passes through the origin, so $W_0 = 0$.
    - $\mathbf{W}^T\mathbf{X} = \mathbf{W} \cdot \mathbf{X} = \|\|\mathbf{W}\|\| \cdot \|\|\mathbf{X}\|\| \cos 90^\circ = 0$
        - This means that the vector $\mathbf{W}$ is perpendicular to the $\mathbf{X}$-axis.
2. **Main Decision Rule for Classification**:
    - $\mathbf{W} \cdot \mathbf{U} + b \geq 0$ for positive classification
    - $\mathbf{W} \cdot \mathbf{U} + b \leq 0$ for negative classification
        - where $\mathbf{W}$ is the hyperplane/decision vector and $\mathbf{U}$ is the sample input vector or data.

- There is a concept called Hard Margin in which wrongly classified data points are not allowed.

3. **Calculating SVM Error**:
   - Margin Error (which is the inverse of $\max f(\mathbf{x})) + C \cdot \text{Classification Error}$
   - $\max f(\mathbf{x}) = \frac{|||\mathbf{W}|||}{2}$
   - Classification Error (ζ) = sum of all ζ values of data points
     - ζ value of correctly classified data = 0
     - ζ value of wrongly classified data = distance between wrongly classified data and the correct support vector
   - C is a hyperparameter; maximizing it means focusing on reducing classification error and classifying all data correctly, while minimizing C means focusing on maximizing the margin.
4. Kernel Trick

   The kernel trick is a technique used in SVMs to handle non-linear data effectively. It involves transforming the data into a higher-dimensional space where it becomes linearly separable.

   1. **Purpose**: The kernel trick is used for classifying data that is in a non-linear form.
   2. **Kernel Functions**: Some commonly used kernel functions in SVMs are:
      - Radial Basis Function (RBF) kernel
      - Polynomial kernel
      - Sigmoid kernel
   3. **Transformation of Non-linear Data**: The kernel trick allows transforming non-linear data into a higher-dimensional space, making it easier to classify.
   4. **Advantages**:
      - Enables SVMs to handle non-linear data
      - Avoids the need to explicitly compute the transformation to the higher-dimensional space
      - Computationally efficient compared to explicitly transforming the data

   The kernel trick is a powerful technique that extends the capabilities of SVMs to handle complex, non-linear data by implicitly mapping it to a higher-dimensional space where it becomes linearly separable.