Summary of Support Vector Machines

Rishabh Iyer

March 19, 2024

Support Vector Machines (SVMs) are a powerful class of supervised learning algorithms for classification. This is the final blog article on SVMs which summarizes the previous three articles. In the first one, we studied SVMs with linearly separable data. Next, we looked at Kernel SVMs. The third article was focused on SVMs with non-separable data.

This summary outlines the steps involved in building and evaluating an SVM model for classification tasks, particularly focusing on binary classification.

1. **Collect the Dataset:** Gather a dataset consisting of input feature vectors (x) and corresponding binary target labels (y), where $y \in \{-1, 1\}$. The dataset should ideally be preprocessed and normalized if necessary.

2. **Choose the Model:** Decide on the SVM model to use based on the data's characteristics:

   - For linearly separable data, use a linear SVM.
   - For non-linearly separable data, select a kernel SVM with an appropriate kernel function (e.g., linear, polynomial, RBF, sigmoid) to implicitly map input features into a higher-dimensional space.

3. **Formulate the Loss Function and Optimization Problem:**

   - In the linear case, minimize the hinge loss with or without regularization to find the optimal separating hyperplane.
   - In the kernel case, use the dual formulation to incorporate the kernel function, focusing on maximizing the margin between classes in the transformed feature space.

4. **Optimize the Model:** Solve the optimization problem using techniques suitable for the chosen formulation:

   - For linear SVMs, gradient descent or similar methods can be used to minimize the loss function.
   - For kernel SVMs, optimization algorithms like Sequential Minimal Optimization (SMO) are employed to solve the dual problem efficiently.

5. **Training Procedure Output:** The training process yields the model parameters (w, b) in the linear case, or the set of support vectors and their corresponding coefficients (α_i) in the kernel case, which define the decision function.

6. **Predict on Test Dataset:** Use the trained SVM model to predict the class labels of new, unseen examples. The decision function, based on the model parameters or support vectors, computes the sign of $w^T x + b$ or the kernelized decision function to classify the inputs.

7. **Evaluate the Model:** Determine the model's performance on a separate test dataset using evaluation metrics such as accuracy, precision, recall, F1 score, and possibly ROC-AUC for binary classification tasks. This step assesses the model's generalization ability and effectiveness in classifying examples.

SVMs offer a robust and versatile approach to classification, capable of handling both linearly and non-linearly separable data with high accuracy. By carefully selecting the model type, kernel function, and optimization strategy, practitioners can leverage SVMs to solve a wide range of classification problems, benefiting from their strong theoretical foundations and generalization capabilities.