The Mechanism of SVM

Support Vector Machines SVM) work by finding the optimal hyperplane that best separates data points of two classes with the maximum margin. The key mechanism and formulation are as follows:

Linear SVM Hyperplane and Margin

- The decision boundary (hyperplane) in the feature space is defined as:

$$w^T x + b = 0$$

Where w is the normal vector perpendicular to the hyperplane, x is a feature vector, and b is the bias term.

- The goal is to find w and b that maximize the margin, which is the distance between the hyperplane and the closest data points (support vectors) of each class.
- The functional margin for each training example (x_i, y_i) with label $y_i \in \{+1, -1\}$ is:

$$y_i(w^T x_i + b) \geq \mathbf{1}$$

This constraint ensures data points are correctly classified and lie outside the margin boundaries.

- The margin width is:

$$\text{Margin} = \frac{2}{||w||}$$

Maximizing margin is equivalent to minimizing $||w||^2/2$.

Primal Optimization Problem

The optimization to find the best hyperplane is:

$$\min_{w,b} \frac{\mathbf{1}}{2}||w||^2 \quad \text{subject to} \quad y_i(w^T x_i + b) \geq \mathbf{1}, \quad \forall i$$

Soft Margin Extension

For non-linearly separable data, slack variables $\xi_i \geq 0$ allow some misclassification with a penalty term C:

$$\min_{w,b,\xi} \frac{\mathbf{1}}{2}||w||^2 + C \sum_i \xi_i$$

subject to

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

Here, C balances margin maximization and misclassification.

Dual Problem and Kernel Trick

Using Lagrange multipliers α_i, the dual form becomes:

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

subject to

$$\sum_i \alpha_i y_i = 0, \quad \alpha_i \geq 0$$

where $K(x_i, x_j)$ is the kernel function enabling non-linear separation by implicitly mapping data to a higher-dimensional space.

Decision Function

Once optimal α_i are found, SVM predicts labels by:

$$f(x) = \text{sign}\left(\sum_i \alpha_i y_i K(x_i, x) + b\right)$$

Only points with $\alpha_i > 0$ are the support vectors influencing the decision boundary.

In summary, SVM finds the maximum margin hyperplane by solving a quadratic optimization problem under constraints, optionally allowing soft margins for non-separable data and extending to non-linear separation with kernels.

If needed, detailed visual explanations or kernel examples can be provided as well.
This explanation references the mathematical derivations and concepts of SVM found in comprehensive articles on SVM mathematics and machine learning theory. [1] [2] [3]

⁂