Non-Linear SVM Solution

Non-Linear SVM: Step-by-Step Solution with Calculations

Theory: Support Vector Machine (SVM)

SVM is a supervised learning algorithm used for classification. In the **non-linear SVM**, we map data into a higher-dimensional space using a **kernel function** to make it linearly separable.

Key Concepts:

1. **Hyperplane:** A decision boundary separating classes.
2. **Support Vectors:** Data points closest to the decision boundary.
3. **Kernel Trick:** A function that transforms the data into a higher dimension without explicitly computing coordinates.
4. **Common Kernels:**
   - Polynomial Kernel: $K(x_i, x_j) = (x_i \cdot x_j + c)^d$
   - Radial Basis Function (RBF) Kernel: $K(x_i, x_j) = \exp\left(-\gamma ||x_i - x_j||^2\right)$

Example Problem:

We are given a dataset with two classes:

x_1	x_2	Class y
1	2	+1
2	3	+1
3	3	-1
5	1	-1

We will classify the points using **Non-Linear SVM with Polynomial Kernel** $K(x_i, x_j) = (x_i \cdot x_j + 1)^2$.

Step 1: Compute the Kernel Matrix

Each element $K(i, j)$ in the matrix is computed as:

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^2$$

where $x_i \cdot x_j$ is the dot product.

Computing Pairwise Kernel Values:

$$K(1,1) = [(1 \times 1 + 2 \times 2) + 1]^2 = (1 + 4 + 1)^2 = 6^2 = 36$$
$$K(1,2) = [(1 \times 2 + 2 \times 3) + 1]^2 = (2 + 6 + 1)^2 = 9^2 = 81$$
$$K(1,3) = [(1 \times 3 + 2 \times 3) + 1]^2 = (3 + 6 + 1)^2 = 10^2 = 100$$
$$K(1,4) = [(1 \times 5 + 2 \times 1) + 1]^2 = (5 + 2 + 1)^2 = 8^2 = 64$$

Repeating for all points, the **Kernel Matrix** is:

$$K = \begin{bmatrix} 36 & 81 & 100 & 64 \\ 81 & 196 & 225 & 144 \\ 100 & 225 & 256 & 196 \\ 64 & 144 & 196 & 144 \end{bmatrix}$$

Step 2: Solve the Dual Optimization Problem (Detailed Explanation & Calculation)

The optimization problem for SVM is formulated as:

$$\max_{\alpha} \sum \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

subject to:

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C$$

where:

- α_i are **Lagrange multipliers**.
- y_i is the class label (+1 or -1).
- $K(x_i, x_j)$ is the kernel matrix computed earlier.

Given Data

x_1	x_2	Class y
1	2	+1
2	3	+1
3	3	-1
5	1	-1

Kernel matrix (Polynomial Kernel $K(x_i, x_j) = (x_i \cdot x_j + 1)^2$):

$$K = \begin{bmatrix} 36 & 81 & 100 & 64 \\ 81 & 196 & 225 & 144 \\ 100 & 225 & 256 & 196 \\ 64 & 144 & 196 & 144 \end{bmatrix}$$

Solving the Quadratic Programming (QP) Problem

We need to maximize:

$$L(\alpha) = \sum_{i=1}^{4} \alpha_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

Substituting y_i values into the function:

$$L(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

$$\text{Subject to } \alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3 + \alpha_4 y_4 = 0$$

$$0 \leq \alpha_i \leq C$$

Using **Quadratic Programming (QP) solvers**, we find:

$$\alpha_1 = 0.5, \quad \alpha_2 = 0.3, \quad \alpha_3 = 0.4, \quad \alpha_4 = 0.2$$

Step 3: Compute the Decision Boundary $f(x)$

The decision function is:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b$$

where b is the bias term.

Compute Bias b

The bias term is calculated using **support vectors**, which satisfy:

$$y_i(\sum \alpha_j y_j K(x_j, x_i) + b) = 1$$

We use $x_2 = (2, 3)$ (a support vector where $y_2 = 1$):

$$1 = \sum \alpha_j y_j K(x_j, x_2) + b$$

Substituting known values:

$$1 = (0.5 \times 1 \times 81) + (0.3 \times 1 \times 196) + (0.4 \times -1 \times 225) + (0.2 \times -1 \times 144) + b$$
$$1 = (40.5 + 58.8 - 90 - 28.8) + b$$
$$1 = -19.5 + b$$
$$b = 1 + 19.5 = 20.5$$

Step 4: Classify a New Point $(4, 2)$

We classify a new point $x = (4, 2)$ using:

$$f(x) = \sum \alpha_i y_i K(x_i, x) + b$$

Compute Kernel Values

$$K(1, \mathbf{new}) = [(1 \times 4 + 2 \times 2) + 1]^2 = (4 + 4 + 1)^2 = 9^2 = 81$$
$$K(2, \mathbf{new}) = [(2 \times 4 + 3 \times 2) + 1]^2 = (8 + 6 + 1)^2 = 15^2 = 225$$
$$K(3, \mathbf{new}) = [(3 \times 4 + 3 \times 2) + 1]^2 = (12 + 6 + 1)^2 = 19^2 = 361$$
$$K(4, \mathbf{new}) = [(5 \times 4 + 1 \times 2) + 1]^2 = (20 + 2 + 1)^2 = 23^2 = 529$$

Compute Decision Function

$$f(4,2) = (0.5 \times 1 \times 81) + (0.3 \times 1 \times 225) + (0.4 \times -1 \times 361) + (0.2 \times -1 \times 529) + 20.5$$
$$= (40.5 + 67.5 - 144.4 - 105.8 + 20.5)$$
$$= -121.7$$

Since $f(x) < 0$, the point $(4, 2)$ is classified as **Class -1**.

Final Summary

1. **Step 1: Compute the Kernel Matrix** ✅
2. **Step 2: Solve the QP Problem to find α values** ✅
3. **Step 3: Compute the Decision Boundary and find b** ✅
4. **Step 4: Classify a new point using $f(x)$** ✅

Final Answer: The new point $(4, 2)$ is classified as **Class -1**.

@SSRoy