Formulas and Numerical Example for Linear Support Vector Machine (SVM)

1 Introduction

This document presents the key mathematical formulas used in Linear Support Vector Machines (SVMs), followed by a numerical example to illustrate the calculations. The formulas define the hyperplane, margin, primal and dual optimization problems, and their constraints.

2 Hyperplane

The separating hyperplane in a Linear SVM is defined by:

$$w^T x + b = 0 \tag{1}$$

where:

- w: Weight vector (normal to the hyperplane).
- x: Feature vector.
- b: Bias term.

3 Margin

The margin is the distance between the two margin hyperplanes ($w^T x + b = 1$ and $w^T x + b = -1$) and is given by:

$$\text{Margin} = \frac{2}{\|w\|}, \quad \text{where} \quad \|w\| = \sqrt{w^T w} \tag{2}$$

4 Primal Optimization Problem

The goal of Linear SVM (hard-margin) is to maximize the margin by minimizing the norm of the weight vector, formulated as:

$$\text{Minimize: } \frac{1}{2}\|w\|^2 \tag{3}$$

Subject to the constraints:

$$y_i(w^T x_i + b) \geq 1, \quad \forall i \tag{4}$$

where:

- y_i: Class label (+1 or −1).
- x_i: Feature vector of the i-th data point.

5 Soft-Margin Optimization Problem

For non-linearly separable data, the soft-margin SVM introduces slack variables ξ_i:

$$\text{Minimize: } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \tag{5}$$

Subject to the constraints:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \tag{6}$$

where C is the regularization parameter controlling the trade-off between margin maximization and classification error.

6 Dual Optimization Problem

The dual form of the SVM optimization problem is:

$$\text{Maximize: } W(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i^T x_j) \tag{7}$$

Subject to the constraints:

$$\sum_{i=1}^{n}\alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C, \quad \forall i \tag{8}$$

where α_i are the Lagrange multipliers, and support vectors have $\alpha_i > 0$.

7 Weight and Bias

The weight vector w is computed as:

$$w = \sum_{i=1}^{n}\alpha_i y_i x_i \tag{9}$$

The bias b is calculated using a support vector ($\alpha_i > 0$) satisfying:

$$y_i(w^T x_i + b) = 1 \quad \Rightarrow \quad b = y_i - w^T x_i \tag{10}$$

8 Prediction

For a new data point x, the class is predicted using:

$$f(x) = w^T x + b \tag{11}$$

- If $f(x) \geq 0$, predict class +1.
- If $f(x) < 0$, predict class −1.

9 Numerical Example

Consider a simple dataset with two 2D data points:

- $x_1 = (1, 1), y_1 = +1$
- $x_2 = (0, 0), y_2 = -1$

9.1 Optimization

Assume the optimization (e.g., using Sequential Minimal Optimization) yields:

$$\alpha_1 = 0.5, \quad \alpha_2 = 0.5 \tag{12}$$

Check the constraint:

$$\alpha_1 y_1 + \alpha_2 y_2 = 0.5 \cdot 1 + 0.5 \cdot (-1) = 0.5 - 0.5 = 0 \tag{13}$$

9.2 Weight Vector

Calculate the weight vector w:

$$w = \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2 = (0.5 \cdot 1 \cdot (1, 1)) + (0.5 \cdot (-1) \cdot (0, 0)) = (0.5, 0.5) + (0, 0) = (0.5, 0.5) \tag{14}$$

9.3 Bias

Using the support vector $x_1 = (1, 1), y_1 = +1$:

$$y_1(w^T x_1 + b) = 1 \tag{15}$$

$$w^T x_1 = (0.5 \cdot 1) + (0.5 \cdot 1) = 0.5 + 0.5 = 1 \tag{16}$$

$$1 \cdot (1 + b) = 1 \quad \Rightarrow \quad b = 0 \tag{17}$$

9.4 Hyperplane

The separating hyperplane is:

$$w^T x + b = 0.5x_1 + 0.5x_2 = 0 \quad \Rightarrow \quad x_1 + x_2 = 0 \tag{18}$$

9.5 Margin

Calculate the margin:

$$\|w\| = \sqrt{(0.5)^2 + (0.5)^2} = \sqrt{0.25 + 0.25} = \sqrt{0.5} \approx 0.707 \tag{19}$$

$$\text{Margin} = \frac{2}{\|w\|} = \frac{2}{\sqrt{0.5}} \approx 2.828 \tag{20}$$

9.6 Prediction

For a new point $x = (2, 2)$:

$$f(x) = w^T x + b = (0.5 \cdot 2) + (0.5 \cdot 2) + 0 = 1 + 1 = 2 \tag{21}$$

Since $f(x) = 2 > 0$, the point is predicted as class $+1$.