Support Vector Machines

G V V Sharma*

Contents

1 **Reflection** 1

2 **Optimization Problem** 1

3 **SVM** 2

***Abstract*—This manual provides an introduction to SVM.**

1 Reflection

1.1 Find the distance of $\mathbf{x}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ from the line

$$\begin{pmatrix} 3 & 4 \end{pmatrix} \mathbf{x} + 5 = 0 \tag{1.1}$$

1.2 Show that the distance of the point $\mathbf{x}_1$ from the line

$$\mathbf{n}^T \mathbf{x} + c = 0, \quad \|\mathbf{n}\| = 1 \tag{1.2}$$

is

$$M = \left|\mathbf{n}^T \mathbf{x}_1 + c\right| \tag{1.3}$$

1.3 Find the reflection $\mathbf{x_2}$ of $\mathbf{x_1}$.

1.4 Define

$$f(\mathbf{x}) = \mathbf{n}^T \mathbf{x} + c \tag{1.4}$$

1.5 Compute $f(\mathbf{x_1})$ and $f(\mathbf{x_2})$. Comment.

2 Optimization Problem

2.1 Suppose $(\mathbf{x_1}, y_1)$ and $(\mathbf{x_2}, y_2)$ where $y_1, y_2 \in \{1, -1\}$. If you want to find $\mathbf{n}, c$ from the given dataset, how will formulate the equivalent optimization problem?

Solution: Consider the optimization problem

$$\max_{\mathbf{n},c} M \tag{2.1}$$

$$\text{s.t} \quad y_1\left(\mathbf{x}_1^T \mathbf{n} + c\right) \geq M \tag{2.2}$$

$$y_2\left(\mathbf{x}_2^T \mathbf{n} + c\right) \geq M \tag{2.3}$$

$$\|\mathbf{n}\| = 1 \tag{2.4}$$

2.2 The *signum* function is defined as

$$\operatorname{sgn}(z) = \begin{cases} 1 & z > 0 \\ 0 & z = 0 \\ -1 & z < 0 \end{cases} \tag{2.5}$$

Show that

$$\operatorname{sgn}(\mathbf{x}^T \mathbf{n} + c) = \operatorname{sgn}(\mathbf{x}^T \mathbf{w} + d) \tag{2.6}$$

where

$$\mathbf{w} = \frac{\mathbf{n}}{M}, d = \frac{c}{M}, M > 0 \tag{2.7}$$

2.3 Show that (2.1)-(2.4) can be reformulated as

$$\min_{\mathbf{w},d} \frac{1}{2}\|\mathbf{w}\|^2 \tag{2.8}$$

$$\text{s.t} \quad y_i\left(\mathbf{x}_i^T \mathbf{w} + d\right) \geq 1 \tag{2.9}$$

Solution: From (2.7),

$$\mathbf{w} = \frac{\mathbf{n}}{M} \implies \|\mathbf{w}\| = \frac{\|\mathbf{n}\|}{M} \tag{2.10}$$

$$\implies M = \frac{1}{\|\mathbf{w}\|} \because \|n\| = 1 \tag{2.11}$$

Thus,

$$\max_{\mathbf{n},c} M = \max_{\mathbf{w},d} \frac{1}{\|\mathbf{w}\|} = \min_{\mathbf{w},d} \|\mathbf{w}\|. \tag{2.12}$$

Also, (2.2)-(2.3) become

$$y_i\left(\mathbf{x}_i^T \mathbf{w} + d\right) \geq 1 \tag{2.13}$$

* The authors are with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India e-mail: gadepall@iith.ac.in.

2.4 Solve (2.8) for $\mathbf{x}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, y_1 = 1$ and $\mathbf{x}_2 = \begin{pmatrix} 0.8 \\ -0.6 \end{pmatrix}, y_2 = -1$ graphically.

2.5 Repeat the above exercise using KKT conditions.

2.6 Repeat the above exercise using *cvxpy/cvxopt*.

3 SVM

3.1 If α_i be the Lagrange multiplier, obtain the Lagrange primal function for (2.8).
Solution: The desired function is given by

$$L_p(\mathbf{w}, \alpha_i) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} \alpha_i \left\{ y_i \left(\mathbf{x}_i^T \mathbf{w} + d - 1 \right) \right\} \tag{3.1}$$

3.2 Show that stationarity condition yields

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{w}_i \tag{3.2}$$

$$0 = \sum_{i=1}^{N} \alpha_i y_i \tag{3.3}$$

Solution: From the stationarity condition,

$$\nabla_{\mathbf{w}} L_p(\mathbf{w}, d, \alpha_i) = \frac{\partial L_p(\mathbf{w}, \alpha_i)}{\partial \mathbf{w}} = 0 \tag{3.4}$$

$$\implies \mathbf{w}^T - \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i^T = 0 \tag{3.5}$$

$$\text{or,} \quad \mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i \tag{3.6}$$

and

$$\nabla_d L_p(\mathbf{w}, d, \alpha_i) = \frac{\partial L_p(\mathbf{w}, \alpha_i)}{\partial d} = 0 \tag{3.7}$$

$$\implies \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{3.8}$$

3.3 Substitute (3.21)-(3.23) in the primal function to obtain the Lagrangian (Wolfe) dual objective function L_D.

3.4 Repeat the above exercises for

$$\min_{\mathbf{w}} \frac{1}{2}||\mathbf{w}||^2 + C \sum_{i=1}^{N} \xi_i \tag{3.9}$$

$$\text{s.t} \quad \xi_i \geq 0 \tag{3.10}$$

$$y_i \mathbf{x}_i^T \mathbf{w} \geq 1 - \xi_i \tag{3.11}$$

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{w}_i \tag{3.12}$$

$$0 = \sum_{i=1}^{N} \alpha_i y_i \tag{3.13}$$

$$\alpha_i = C - \mu_i \tag{3.14}$$

3.5 Show that the KKT conditions are

$$\alpha_i \left[y_i \left(\mathbf{x}_i^T \mathbf{w} \right) - (1 - \xi_i) \right] = 0 \tag{3.15}$$

$$\mu_i \xi_i = 0 \tag{3.16}$$

$$y_i \left(\mathbf{x}_i^T \mathbf{w} \right) - (1 - \xi_i) = 0 \tag{3.17}$$

3.6

$$P = \left\{ \mathbf{x} : f(\mathbf{x}) = \mathbf{n}^T \mathbf{x} + c = 0 \right\}, \quad ||\mathbf{n}|| = 1 \tag{3.18}$$

be a hyperplane where $\mathbf{n}$ is a unit normal vector to the plane.

3.7 Let

$$P = \left\{ \mathbf{x} : f(\mathbf{x}) = \mathbf{n}^T \mathbf{x} + c = 0 \right\}, \quad ||\mathbf{n}|| = 1 \tag{3.19}$$

be a hyperplane where $\mathbf{n}$ is a unit normal vector to the plane.

3.8 Consider the quadratic programming problem

$$\min_{\mathbf{w}} \frac{1}{2}||\mathbf{w}||^2 + C \sum_{i=1}^{N} \xi_i \tag{3.20}$$

$$\text{s.t } \xi_i \geq 0 \tag{3.21}$$

$$y_i \left(\mathbf{x}_i^T \mathbf{w} \right) \geq 1 - \xi_i \tag{3.22}$$

3.9 If α_i, μ_i be the Lagrange multipliers, obtain the Lagrange primal function.

3.10 Show that stationarity condition yields

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{w}_i \tag{3.23}$$

$$0 = \sum_{i=1}^{N} \alpha_i y_i \tag{3.24}$$

$$\alpha_i = C - \mu_i \tag{3.25}$$

3.11 Show that the KKT conditions are

$$\alpha_i \left[y_i \left(\mathbf{x}_i^T \mathbf{w} \right) - (1 - \xi_i) \right] = 0 \tag{3.26}$$

$$\mu_i \xi_i = 0 \tag{3.27}$$

$$y_i \left(\mathbf{x}_i^T \mathbf{w} \right) - (1 - \xi_i) = 0 \tag{3.28}$$

3.12 Substitute (3.21)-(3.23) in the primal function to obtain the Lagrangian (Wolfe) dual objective function L_D.