

G V V Sharma\*

## CONTENTS

1	Reflection	1
2	Optimization	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

$$(3 \ 4)\mathbf{x} + 5 = 0 \quad (1.1)$$

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

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

is

$$M = |\mathbf{n}^T \mathbf{x}_1 + c| \quad (1.3)$$

1.3 Find the reflection  $\mathbf{x}_2$  of  $\mathbf{x}_1$ .

1.4 Define

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

1.5 Compute  $f(\mathbf{x}_1)$  and  $f(\mathbf{x}_2)$ . Comment.

### 2 OPTIMIZATION

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 op-

timization problem?

**Solution:** Consider the optimization problem

$$\max_{\mathbf{n}, c} M \quad (2.1)$$

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

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

$$\|\mathbf{n}\| = 1 \quad (2.4)$$

2.2 The *signum* function is defined as

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

Show that

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

where

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

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

$$\min_{\mathbf{w}, d} \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.8)$$

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

**Solution:** From (2.7),

$$\mathbf{w} = \frac{\mathbf{n}}{M} \implies \|\mathbf{w}\| = \frac{\|\mathbf{n}\|}{M} \quad (2.10)$$

$$\implies M = \frac{1}{\|\mathbf{w}\|} \because \|\mathbf{n}\| = 1 \quad (2.11)$$

Thus,

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

Also, (2.2)-(2.3) become

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

2.4 Solve (2.8) using *cvxpy/cvxopt* for  $\mathbf{x}_1 =$

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

$\begin{pmatrix} 2 \\ 1 \end{pmatrix}, y_1 = 1$  and  $\mathbf{x}_2 = \begin{pmatrix} 0.8 \\ -0.6 \end{pmatrix}, y_2 = -1$ .

**Solution:** From the given information, the constraints in (2.8) become

$$\begin{pmatrix} 2 & 1 \end{pmatrix} \mathbf{w} + d \geq 1 \quad (2.14)$$

$$\begin{pmatrix} 0.8 & -0.6 \end{pmatrix} \mathbf{w} + d \leq -1 \quad (2.15)$$

The following code results in

$$\mathbf{w}_{opt} = \begin{pmatrix} 0.6 & 0.8 \end{pmatrix}, d_{opt} = 1, \|\mathbf{w}_{opt}\|^2 = 1 \quad (2.16)$$

```
import cvxpy as cp

w = cp.Variable(2)
d = cp.Variable()

probconst = ([2*w[0]+w[1]+d>=1,0.8*w
[0]-0.6*w[1]+d<=-1])
probojb = cp.Minimize(0.5*cp.square(cp.
norm(w)))

prob = cp.Problem(probojb,probconst)
prob.solve()

print (prob.value)
print (w.value)
print (d.value)
```

2.5 Solve (2.8) graphically.

**Solution:** The following code plots Fig. 2.5. The constraint lines in (2.14)-(2.15) are plotted for  $d = 0, 0.5$  and  $1$ . The circles  $\|\mathbf{w}\|^2 = r^2$  are plotted for  $r = 1, 2$  and  $3$ . The smallest circle that satisfies the constraints is obtained when  $d = 1$

```
wget https://raw.githubusercontent.com/
gadepall/EE1390/master/manuals/svm/
codes/svm_graph.py
```

2.6 Repeat the above exercise using KKT conditions.

### 3 SVM

3.1 If  $\alpha_i$  be the Lagrange multiplier, obtain the Lagrange primal function for (2.8).

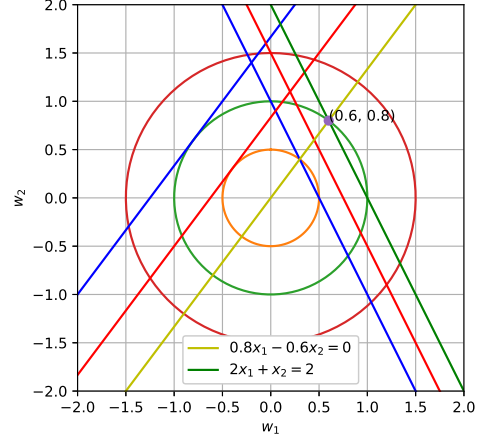


Fig. 2.5

**Solution:** The desired function is given by

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

3.2 Show that stationarity condition yields

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

$$0 = \sum_{i=1}^N \alpha_i y_i \quad (3.3)$$

**Solution:** From the stationarity condition,

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

$$\Rightarrow \mathbf{w}^T - \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i^T = 0 \quad (3.5)$$

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

and

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

$$\Rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \quad (3.8)$$

3.3 Substitute (3.2)-(3.3) in the primal function to obtain the Lagrangian (Wolfe) dual objective function  $L_D$ .

**Solution:** Substituting (3.2) in (3.1),

$$L_D(\alpha_i) = \frac{1}{2} \left\| \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \right\|^2 - \sum_{i=1}^N \alpha_i \left\{ y_i \left( \mathbf{x}_i^T \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i + d - 1 \right) \right\} \quad (3.9)$$

in

3.4 Repeat the above exercises for

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (3.10)$$

$$\text{s.t. } \xi_i \geq 0 \quad (3.11)$$

$$y_i \mathbf{x}_i^T \mathbf{w} \geq 1 - \xi_i \quad (3.12)$$

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{w}_i \quad (3.13)$$

$$0 = \sum_{i=1}^N \alpha_i y_i \quad (3.14)$$

$$\alpha_i = C - \mu_i \quad (3.15)$$

3.5 Show that the KKT conditions are

$$\alpha_i \left[ y_i (\mathbf{x}_i^T \mathbf{w}) - (1 - \xi_i) \right] = 0 \quad (3.16)$$

$$\mu_i \xi_i = 0 \quad (3.17)$$

$$y_i (\mathbf{x}_i^T \mathbf{w}) - (1 - \xi_i) = 0 \quad (3.18)$$

3.6

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

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

3.7 Let

$$P = \left\{ \mathbf{x} : f(\mathbf{x}) = \mathbf{n}^T \mathbf{x} + c = 0 \right\}, \quad \|\mathbf{n}\| = 1 \quad (3.20)$$

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

3.8 Consider the quadratic programming problem

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (3.21)$$

$$\text{s.t. } \xi_i \geq 0 \quad (3.22)$$

$$y_i (\mathbf{x}_i^T \mathbf{w}) \geq 1 - \xi_i \quad (3.23)$$

3.9 If  $\alpha_i, \mu_i$  be the Lagrange multipliers, obtain the Lagrange primal function.

3.10 Show that stationarity condition yields

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{w}_i \quad (3.24)$$

$$0 = \sum_{i=1}^N \alpha_i y_i \quad (3.25)$$

$$\alpha_i = C - \mu_i \quad (3.26)$$

3.11 Show that the KKT conditions are

$$\alpha_i \left[ y_i (\mathbf{x}_i^T \mathbf{w}) - (1 - \xi_i) \right] = 0 \quad (3.27)$$

$$\mu_i \xi_i = 0 \quad (3.28)$$

$$y_i (\mathbf{x}_i^T \mathbf{w}) - (1 - \xi_i) = 0 \quad (3.29)$$

3.12 Substitute (3.2)-(3.3) in the primal function to obtain the Lagrangian (Wolfe) dual objective function  $L_D$ .