# Support Vector Machines

G V V Sharma*

Contents



***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

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

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

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

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

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

2.2 The *signum* function is defined as

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

Show that

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

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

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

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

**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

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

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

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

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

The following code results in

$$\mathbf{w} = \begin{pmatrix}0.6 & 0.8\end{pmatrix}, d_{opt} = 1 \tag{2.16}$$

```
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])
probobj = cp.Minimize(0.5*cp.square(cp.
     norm(w)))

prob = cp.Problem(probobj,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

```
import numpy as np
import matplotlib.pyplot as plt
#If using termux
import subprocess
import shlex
#end if

x = np.zeros(4)
y = np.zeros(4)
r = np.arange(4)
phi = np.linspace(0.0,2*np.pi,100)
na=np.newaxis

x_line = x[na,:]+r[na,:]*np.sin(phi[:,na])/2
y_line = y[na,:]+r[na,:]*np.cos(phi[:,na])/2
ax=plt.plot(x_line,y_line)
w1 = np.linspace(−2,2,100)

d = 0
w21 = (0.8*w1 + 1 + d)/0.6
w22 = (1−d−2*w1)
plt.plot(w1,w21,color = 'b')
plt.plot(w1,w22,color = 'b')

d = −0.5
w21 = (0.8*w1 + 1 + d)/0.6
w22 = (1−d−2*w1)
plt.plot(w1,w21,color = 'r')
plt.plot(w1,w22,color = 'r')

d = −1
w21 = (0.8*w1 + 1 + d)/0.6
w22 = (1−d−2*w1)
plt.plot(w1,w21,color = 'y',label = '$0.8x_1
     − 0.6x_2 = 0$')
plt.plot(w1,w22,color = 'g',label = '$2x_1 +
      x_2 = 2$')

plt.axis([−2, 2, −2, 2])
plt.gca().set_aspect('equal', adjustable='box')
plt.grid()
plt.xlabel('$w_1$')
plt.ylabel('$w_2$')
plt.legend(loc='best')

A = [0.6]
B = [0.8]

plt.plot(A,B,'o')
for xy in zip(A, B):
    plt.annotate('(%.4s, %.4s)' % xy, xy=xy,
         textcoords='data')
#If using termux
plt.savefig('../figs/svm_graph.pdf')
plt.savefig('../figs/svm_graph.eps')
subprocess.run(shlex.split("termux−open ../
    figs/svm_graph.pdf"))
#else
#plt.show()
```

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).

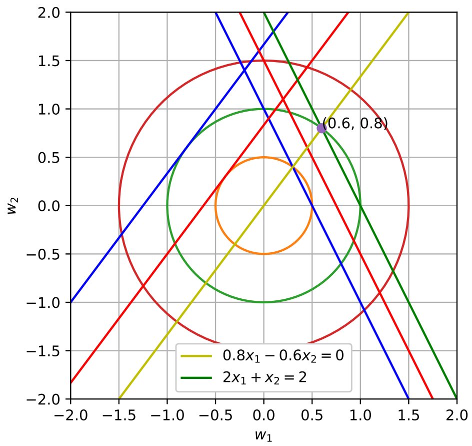


Fig. 2.5

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

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

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

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

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

in

3.4 Repeat the above exercises for

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

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

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

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

$$0 = \sum_{i=1}^{N} \alpha_i y_i \tag{3.14}$$

$$\alpha_i = C - \mu_i \tag{3.15}$$

3.5 Show that the KKT conditions are

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

$$\mu_i \xi_i = 0 \tag{3.17}$$

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

3.6

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

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

3.7 Let

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

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

3.8 Consider the quadratic programming problem

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

$$\text{s.t} \ \xi_i \geq 0 \tag{3.22}$$

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

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

3.10 Show that stationarity condition yields

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

$$0 = \sum_{i=1}^{N} \alpha_i y_i \tag{3.25}$$

$$\alpha_i = C - \mu_i \tag{3.26}$$

3.11 Show that the KKT conditions are

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

$$\mu_i \xi_i = 0 \tag{3.28}$$

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

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