

G V V Sharma*

CONTENTS

1	Reflection	1
2	Optimization Problem	1
3	Solver	2
4	KKT Solution	2
5	Duality	3
6	Convexification	3

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 PROBLEM

2.1 Suppose (\mathbf{x}_1, y_1) and (\mathbf{x}_2, y_2) are i/o data for a system where $y_1, y_2 \in \{1, -1\}$. If you want to find \mathbf{n}, c from the given dataset, how will you formulate the equivalent optimization 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,

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

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

Also, (2.2)-(2.3) become

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

3 SOLVER

3.1 Solve (2.8) using *cvxpy/cvxopt* for $\mathbf{x}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $y_1 = 1$ and $\mathbf{x}_2 = \begin{pmatrix} 0.8 \\ -0.6 \end{pmatrix}$, $y_2 = -1$.

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

$$\begin{pmatrix} 2 & 1 \end{pmatrix} \mathbf{w} + d \geq 1 \quad (3.1)$$

$$\begin{pmatrix} 0.8 & -0.6 \end{pmatrix} \mathbf{w} + d \leq -1 \quad (3.2)$$

The following code results in

$$\mathbf{w}_{opt} = (0.6 \ 0.8), d_{opt} = 1, \|\mathbf{w}_{opt}\|^2 = 1 \quad (3.3)$$

```
from cvxpy import *
import numpy as np

# x is array of datapoints stacked column
# wise [x1, x2, ... , xn]
x = np.array([
    [ 2.0, 0.8 ],
    [ 1.0, -0.6 ]
])

# y is matrix with labels stacked diagonally
y = np.diag([1,-1])

n=2 #no of datapoints
p=2 #no of parameters of each datapoint

d = Variable()
d_v = np.ones((p,1))*d
#broadcasted d into a vector d_v

w = Variable((p,1),nonneg=False)

#objective function
obj = Minimize(0.5*sum(norm(w)))

#constraints
constraints = [((x@y.T)@w + y@d_v)>= np
    .ones((n,1))]

Problem(obj, constraints).solve()
print("Minimum value of Cost function= ",
    obj.value)
print("Minima is at w = \n",w.value)
```

```
print("Minima is at d= ",d.value)
```

3.2 Provide a graphical representation for (2.8)

Solution: The following code plots Fig. 3.2. The constraint lines in (3.1)-(3.2) 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
```

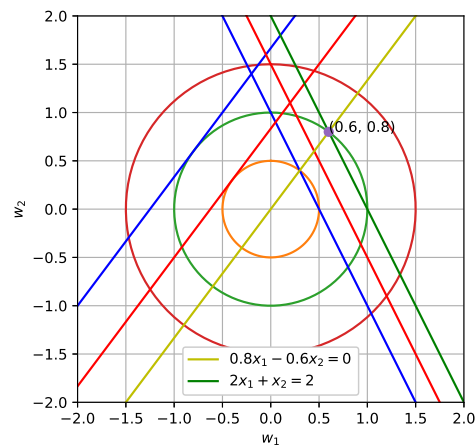


Fig. 3.2

4 KKT SOLUTION

4.1 Show that the Lagrangian for (2.8) can be expressed as

$$L_p(\mathbf{w}, \alpha, d) = \frac{1}{2} \|\mathbf{w}\|^2 - \alpha^T \left(\begin{pmatrix} y_1 \mathbf{x}_1 & y_2 \mathbf{x}_2 \end{pmatrix}^T \mathbf{w} + d \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \quad (4.1)$$

where

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad (4.2)$$

are the Lagrange multipliers.

Solution: The Lagrangian is given by,

$$L_p(\mathbf{w}, \alpha, d) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^2 \alpha_i \{y_i (\mathbf{x}_i^T \mathbf{w} + d) - 1\} \quad (4.3)$$

which can be simplified to obtain (4.1)

- 4.2 Show that the stationarity condition with respect to \mathbf{w} yields

$$\begin{pmatrix} \mathbf{I} & -(y_1 \mathbf{x}_1 & y_2 \mathbf{x}_2) & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \alpha \\ d \end{pmatrix} = \mathbf{0} \quad (4.4)$$

Solution: From the stationarity condition

$$\nabla_{\mathbf{w}} L_p(\mathbf{w}, \alpha, d) = \mathbf{0} \quad (4.5)$$

$$\text{or, } \mathbf{w} - (y_1 \mathbf{x}_1 \quad y_2 \mathbf{x}_2) \alpha = \mathbf{0} \quad (4.6)$$

resulting in (4.4).

- 4.3 Show that the stationarity condition with respect to α yields

$$\begin{pmatrix} (y_1 \mathbf{x}_1 & y_2 \mathbf{x}_2)^T & \mathbf{0} & \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \alpha \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (4.7)$$

Solution:

$$\nabla_{\alpha} L_p(\mathbf{w}, \alpha, d) = 0 \quad (4.8)$$

$$\Rightarrow (y_1 \mathbf{x}_1 \quad y_2 \mathbf{x}_2)^T \mathbf{w} + d \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0 \quad (4.9)$$

after simplification resulting in (4.7)

- 4.4 Find the stationarity condition with respect to d .

Solution:

$$\nabla_d L_p(\mathbf{w}, \alpha, d) = 0 \quad (4.10)$$

$$\Rightarrow (y_1 \quad y_2) \alpha = 0 \quad (4.11)$$

$$\text{or, } \begin{pmatrix} \mathbf{0} & (y_1 \quad y_2) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \alpha \\ d \end{pmatrix} = 0 \quad (4.12)$$

- 4.5 Obtain a matrix equation for \mathbf{w} and d .

Solution: (4.4) (4.7) and (4.12) can be stacked into a single matrix equation as

$$\begin{pmatrix} \mathbf{I} & -(y_1 \mathbf{x}_1 & y_2 \mathbf{x}_2) & \mathbf{0} \\ (y_1 \mathbf{x}_1 & y_2 \mathbf{x}_2)^T & \mathbf{0} & \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ \mathbf{0} & (y_1 \quad y_2) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \alpha \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad (4.13)$$

- 4.6 Find the optimal values of \mathbf{w} and d by solving (4.13) using python.

Solution:

wget https://github.com/gadepall/EE1390/raw/master/manuals/svm/codes/svm_matrix.py

5 DUALITY

- 5.1 Substitute (4.4) and (4.12) in the primal function to obtain the Lagrangian (Wolfe) dual objective function L_D .

Solution: From (4.4)

$$\mathbf{w} = (y_1 \mathbf{x}_1 \quad y_2 \mathbf{x}_2) \alpha \quad (5.1)$$

$$\Rightarrow \mathbf{w}^T \mathbf{w} = \alpha^T \begin{pmatrix} y_1 \mathbf{x}_1^T \\ y_2 \mathbf{x}_2^T \end{pmatrix} (y_1 \mathbf{x}_1 \quad y_2 \mathbf{x}_2) \alpha \quad (5.2)$$

$$\text{and } \alpha^T (y_1 \mathbf{x}_1 \quad y_2 \mathbf{x}_2)^T \mathbf{w} = \mathbf{w}^T \mathbf{w} \quad (5.3)$$

From (4.12),

$$(y_1 \quad y_2) \alpha = 0 \quad (5.4)$$

$$\Rightarrow \alpha^T d \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 0 \quad (5.5)$$

Substituting from the above in (4.1),

$$L_D(\alpha) = -\frac{1}{2} \alpha^T \begin{pmatrix} y_1 \mathbf{x}_1^T \\ y_2 \mathbf{x}_2^T \end{pmatrix} (y_1 \mathbf{x}_1 \quad y_2 \mathbf{x}_2) \alpha + \alpha^T \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (5.6)$$

- 5.2 From (4.1), show that

$$L_p(\mathbf{w}, \alpha, d) \leq \frac{1}{2} \|\mathbf{w}\|^2 \quad (5.7)$$

- 5.3 Let f^* be the solution of (2.8). Show that

$$\min_{\mathbf{w}, d} L_p(\mathbf{w}, \alpha, d) \leq f^* \quad (5.8)$$

- 5.4 Show that

$$L_D(\alpha) = \min_{\mathbf{w}, d} L_p(\mathbf{w}, \alpha, d) \quad (5.9)$$

- 5.5 Show that

$$f^* = \max_{\alpha} L_D(\alpha) \quad (5.10)$$

$$\text{s.t. } \alpha \geq 0 \quad (5.11)$$

$$\nabla_d L_p(\mathbf{w}, \alpha, d) = 0 \quad (5.12)$$

Solution:

wget https://github.com/gadepall/EE1390/raw/master/manuals/svm/codes/wolfe_dual.py

- 5.6 Verify the above result by theoretically solving (5.10).

6 CONVEXIFICATION

- 6.1 Given $(\mathbf{x}_i, y_i), i = 1 \dots n$ are i/o data for a system where $y_i \notin \{1, -1\}$. If you want to

find \mathbf{w}, d from the given dataset, how will you formulate the equivalent optimization problem by modifying (2.8)?

Solution: The desired expression is

$$\min_{\mathbf{w}, d} \frac{1}{2} \|\mathbf{w}\|^2 \quad (6.1)$$

$$\text{s.t. } y_i (\mathbf{x}_i^T \mathbf{w} + d) \geq 1 - \xi_i \quad (6.2)$$

$$\xi_i \geq 0 \quad (6.3)$$

$$\sum_i^n \xi_i \leq C \quad (6.4)$$

Note that C is also a parameter that needs to be given for estimating the parameters \mathbf{w}, d of the SVM engine.

6.2 How will you classify the output y for a given input \mathbf{x} .

Solution:

$$y = \text{sgn}(\mathbf{x}^T \mathbf{w} + d) \quad (6.5)$$