QP Solution to Hard SVM

Dr. Fayyaz ul Amir Afsar Minhas

PIEAS Biomedical Informatics Research Lab

Department of Computer and Information Sciences

Pakistan Institute of Engineering & Applied Sciences

PO Nilore, Islamabad, Pakistan

http://faculty.pieas.edu.pk/fayyaz/

SVM as an Optimization problem

- We can present SVM as the following optimization problem

$$\max \quad \rho = \frac{2}{\|\mathbf{w}\|}$$

$s.t.$

$$\mathbf{w}^T\mathbf{x}^{(i)} + b \le -1, \qquad \forall i \ \mathbf{s.t.}\ y_i = -1$$

$$\mathbf{w}^T\mathbf{x}^{(i)} + b \ge +1, \qquad \forall i \ \mathbf{s.t.}\ y_i = +1$$

OR

$$\min \quad \rho = \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

$s.t.$

$$y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right) \ge +1$$

SVM as an Optimization problem

- Combining the objective function and the constraints (represented as losses) as follows

$$\min_{\mathbf{w},b} \quad M = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{N} L_i$$

Produces a large +ive value when the i[th] constraint is violated

- The constraint cost function can be written as

$$L_i = \max_{\alpha_i} \quad \alpha_i \left\{1 - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right)\right\} = \begin{cases} 0 & 1 - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right) \leq 0 \\ \infty & else \end{cases}$$

$$\alpha_i \geq 0$$

When the constraint is being strictly satisfied, α_i must be zero

α_i must be positive when the constraint is being violated

SVM as an Optimization problem

- Combining

$$Let \qquad P = \min_{\mathbf{w},b}\left\{\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{N}\left[\max_{\alpha_i \geq 0} \alpha_i \left\{1 - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right)\right\}\right]\right\}$$

$$= \min_{\mathbf{w},b}\max_{\alpha_i \geq 0}\underbrace{\left\{\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{N}\left[\alpha_i \left\{1 - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right)\right\}\right]\right\}}_{}$$

This is the Lagrangian (α_i are the Lagrange Multipliers)

- This is called the primal form of the optimization problem
- The corresponding dual can be written as

$$Let \qquad D = \max_{\alpha_i \geq 0}\min_{\mathbf{w},b}\left\{\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{N}\left[\alpha_i \left\{1 - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right)\right\}\right]\right\}$$

SVM as an Optimization problem

- Since the optimization is convex therefore if the Karush-Kuhn-Tucker conditions are satisfied then the primal and dual optimal values will be equal
- In simple words the optimization problem in SVM can be interpreted as

$$\max_{\alpha_i \geq 0} \min_{\mathbf{w},b} \left\{ \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{N}\left[\alpha_i \left\{1 - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right)\right\}\right]\right\}$$

$$s.t. \qquad \alpha_i \geq 0$$

SVM as an Optimization problem

$$\max_{\alpha_i \geq 0} \min_{\mathbf{w},b} \left\{ \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{N} \left[\alpha_i \left\{ 1 - y_i \left(\mathbf{w}^T\mathbf{x}^{(i)} + b \right) \right\} \right] \right\} = \max_{\alpha_i \geq 0} \left\{ \theta_D \left(\mathbf{w}, b \right) \right\}$$

$$s.t. \qquad \alpha_i \geq 0$$

$$\theta_D \left(\mathbf{w}, b \right) = \min_{\mathbf{w},b} \left\{ \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{N} \left[\alpha_i \left\{ 1 - y_i \left(\mathbf{w}^T\mathbf{x}^{(i)} + b \right) \right\} \right] \right\} \qquad [1]$$

Finding the saddle point of $\theta_D \left(\mathbf{w}, b \right)$

$$\frac{\partial \theta_D \left(\mathbf{w}, b \right)}{\partial \mathbf{w}} = \mathbf{w} + \sum_{i=1}^{N} \left[\alpha_i \left\{ -y_i \mathbf{x}^{(i)} \right\} \right] = 0 \qquad \Rightarrow \mathbf{w}^* = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}^{(i)}$$

$$\frac{\partial \theta_D \left(\mathbf{w}, b \right)}{\partial b} = -\sum_{i=1}^{N} \alpha_i y_i = 0 \qquad \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0$$

SVM as an Optimization problem

Putting in [1]

$$\theta_D(\mathbf{w}^*, b^*) = \frac{1}{2}\mathbf{w}^{*T}\mathbf{w}^* + \sum_{i=1}^{N}\left[\alpha_i\left\{1 - y_i\left(\mathbf{w}^{*T}\mathbf{x}^{(i)} + b^*\right)\right\}\right]$$

$$= \frac{1}{2}\left(\sum_{i=1}^{N}\alpha_i y_i \mathbf{x}^{(i)}\right)^T\left(\sum_{j=1}^{N}\alpha_j y_j \mathbf{x}^{(j)}\right) + \sum_{i=1}^{N}\left[\alpha_i\left\{1 - y_i\left(\left(\sum_{j=1}^{N}\alpha_j y_j \mathbf{x}^{(j)}\right)^T\mathbf{x}^{(i)} + b^*\right)\right\}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i y_i \mathbf{x}^{(i)^T}\alpha_j y_j \mathbf{x}^{(j)} + \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\left[\alpha_i y_i\left(\sum_{j=1}^{N}\alpha_j y_j \mathbf{x}^{(j)}\right)^T\mathbf{x}^{(i)} + b^*\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \mathbf{x}^{(i)^T}\mathbf{x}^{(j)} + \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \mathbf{x}^{(i)^T}\mathbf{x}^{(j)} + b^*\sum_{i=1}^{N}\alpha_i y_i$$

$$= \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \mathbf{x}^{(i)^T}\mathbf{x}^{(j)}$$

SVM as an Optimization problem

- Thus the problem can be written as

$$\max_{\alpha_i \geq 0} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^T} \mathbf{x}^{(j)} \right\}$$

$$s.t \qquad \alpha_i \geq 0$$

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

- This quadratic optimization problem can be solved for α_i using standard optimization packages

SVM as an Optimization problem

- The training points with their corresponding α_i greater than zero lie on the boundary and are thus called support vectors as they support the boundary
- Once α_i have been found, the weight can be calculated as

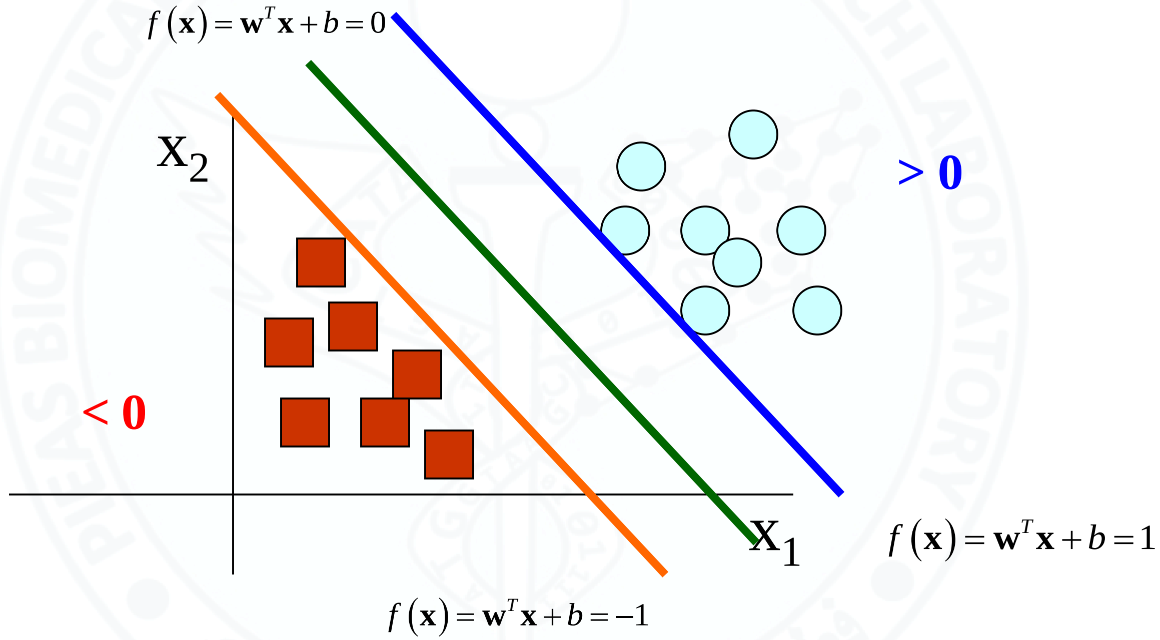
$$\mathbf{w}^* = \sum_{i=1}^{N} \alpha_i^* y_i \mathbf{x}^{(i)}$$

$$\mathbf{w}^{*^T}\mathbf{x} + b^* = y \quad \Rightarrow \quad b^* = y - \mathbf{w}^{*^T}\mathbf{x}$$

$$or \qquad b^* = \frac{1}{n_{SV}} \sum_{\alpha_i > 0} \left(y_i - \mathbf{w}^{*^T}\mathbf{x}^{(i)} \right)$$

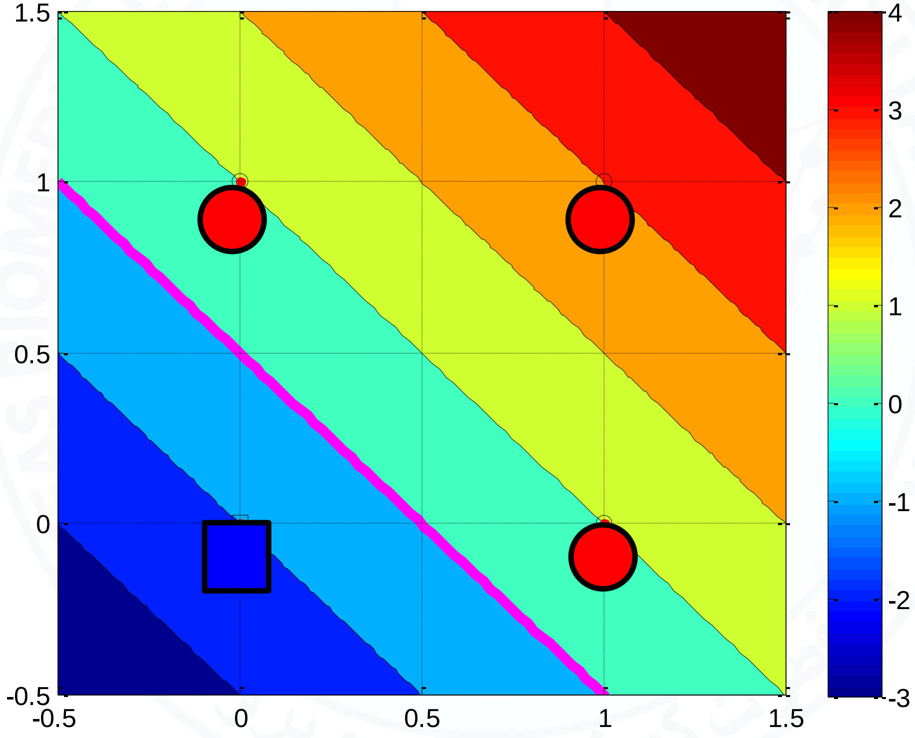
- Classification
  - The label of an unknown point can be determined by

$$y = \mathbf{sgn}\left(w^{*^T} x + b^* \right)$$



$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0$
X_2
> 0
< 0
X_1
$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 1$
$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = -1$

Example problem



Matrix formulation of SVM Problem

$$\max_{\alpha_i \geq 0} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^T} \mathbf{x}^{(j)} \right\}$$

$$s.t \quad \alpha_i \geq 0$$

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

$$max_{\boldsymbol{\alpha}}\ \mathbf{1}^T \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\alpha}$$

Subject to:

$$\boldsymbol{\alpha} \succcurlyeq \mathbf{0}$$

$$\boldsymbol{y}^T \boldsymbol{\alpha} = 0$$

- If we define the following:
  - $\boldsymbol{\alpha} = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_N]^{\boldsymbol{T}}$
  - $\boldsymbol{y} = [y_1 \quad y_2 \quad \cdots \quad y_N]^{\boldsymbol{T}}$
  - $\mathbf{1}_N = [1 \quad 1 \quad \ldots \quad 1]^{\boldsymbol{T}}$
  - $\boldsymbol{X}_{(d \times N)} = [\boldsymbol{x}_{\mathbf{1}} y_1 \quad \boldsymbol{x}_{\mathbf{2}} y_2 \quad \cdots \quad \boldsymbol{x}_N y_N]$

Solving the SVM using QP

- CVXOPT is a Python package that implements a quadratic programming solver
  - http://abel.ee.ucla.edu/cvxopt/userguide/coneprog.html#quadratic-programming
- cvxopt.solvers.qp(*P*, *q*[, *R*, *s*[, *U*, *v*[, *solver*[, *initvals*]]]])
  - Solves the following problem for $\boldsymbol{z}$:

$$min_{\boldsymbol{z}} \frac{1}{2}\boldsymbol{z}^T\boldsymbol{P}\boldsymbol{z} + \boldsymbol{q}^T\boldsymbol{z}$$

Subject to:

$$\boldsymbol{R}\boldsymbol{z} \preccurlyeq \boldsymbol{s}$$
$$\boldsymbol{U}\boldsymbol{z} = \boldsymbol{v}$$

←

$$\boldsymbol{z} = \boldsymbol{\alpha}$$
$$\boldsymbol{P} = \boldsymbol{X}^T\boldsymbol{X}$$
$$\boldsymbol{q} = -\boldsymbol{1}_N$$
$$\boldsymbol{R} = -\boldsymbol{I}_{N\times N}$$
$$\boldsymbol{s} = \boldsymbol{0}_N$$
$$\boldsymbol{U} = \boldsymbol{y}^T$$
$$\boldsymbol{v} = 0$$

SVM Problem

$$max_{\boldsymbol{\alpha}}\ \boldsymbol{1}^T\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\alpha}$$

Subject to:

$$\boldsymbol{\alpha} \succcurlyeq \boldsymbol{0}$$
$$\boldsymbol{y}^T\boldsymbol{\alpha} = 0$$

Programming the SVM

End of Lecture-1

We want to make a machine that will be proud of us.

- Danny Hillis