Convex Optimization

Lab 8: Solve SVM via Quadratic Optimization

Lecturer: *Dr*. Wan-Lei Zhao

Autumn Semester 2024

Contact: wlzhao@xmu.edu.cn

Outline

Overview of discriminative classifer (1)

- Given a training set $(x_i, y_i)_{i=1\cdots m}$
- $x_i \in R^n$ is the observation, $y_i \in [-1, 1]$ is the class label
- A classifier is trained with this set
- Given a new instance $u \in R^n$
- The classifier makes the prediction whether it is -1 or 1

SVM: the model (1)

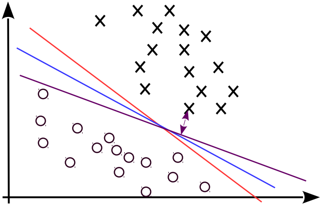


Figure: Searching for maximum γ

$$\gamma = \underset{i=1,\cdots,m}{\text{argmin}}\ \gamma^{(i)} \tag{1}$$

- Searching for w, b that maximizes γ

$$\begin{aligned} \underset{w,b,\gamma}{\text{Max.}} \quad & \gamma \\ \text{s. t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq \gamma, i = 1, \cdots, m \\ & ||w|| = 1. \end{aligned}$$

SVM: the model (2)

- Searching for w, b that maximizes γ

$$\begin{aligned} \underset{w,b,\gamma}{\text{Max.}} \quad & \gamma \\ \text{s. t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq \gamma, i = 1, \cdots, m \\ & ||w|| = 1. \end{aligned} \tag{2}$$

- Unfortunately, above problem is not solvable
- Constraint $||w|| = 1$ is not convex

$$\begin{aligned} \underset{w,b,\hat{\gamma}}{\text{Max.}} \quad & \hat{\gamma}/||w|| \\ \text{s. t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, i = 1, \cdots, m. \end{aligned} \tag{3}$$

- Unfortunately, above problem is not solvable either
- $\hat{\gamma}$ is functional margin, it is valid to scale it to $\hat{\gamma} = 1$

SVM: the model (3)

$$\max_{w,b,\hat{\gamma}} \quad \hat{\gamma}/||w|| \tag{3}$$
$$\text{s. t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq \hat{\gamma}, i = 1, \cdots, m.$$

- $\hat{\gamma}$ is functional margin, it is valid to scale it to $\hat{\gamma} = 1$

$$\max_{w,b} \quad \frac{1}{||w||} \tag{4}$$
$$\text{s. t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, i = 1, \cdots, m.$$

- This is equivalent to solving following **quadratic optimization** problem

$$\min_{w,b} \quad \frac{1}{2}||w||^2 \tag{5}$$
$$\text{s. t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1, i = 1, \cdots, m.$$

SVM: the model (4)

- Binary classification problem is now modeled as **quadratic optimization** problem

$$\begin{aligned} \underset{w,b}{\text{Min.}} \quad & \frac{1}{2}||w||^2 \\ \text{s. t.} \quad & 1 - y^{(i)}(w^T x^{(i)} + b) \leq 0, i = 1, \cdots, m. \end{aligned} \tag{6}$$

- The unknowns are w and **b**
- The target function is quadratic
- m constraints are linear
- In the lecture, we solve it with **Lagrange multiplier** method
- Here we solve it with **quadratic programming (QP)**

Outline

Quadratic Programming: standard form (1)

- Given the model for a problem:

$$\text{Min. } 2x_1^2 + x_2^2 - x_1x_2 - 4x_1 - 3x_2$$

$$\text{s. t.} \begin{cases} x_1 + x_2 \leq 4 \\ -x_1 + 3x_2 \leq 3 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

- Standard Quadratic Optimization form:

$$\begin{aligned} \text{Min. } & \frac{1}{2}x^T Hx + c^T x \\ \text{s. t.} & Ax \preccurlyeq b \\ & Aeqx = Beq \\ & lb \preccurlyeq x \preccurlyeq ub \end{aligned}$$

$$H = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}, c = [\text{-4 -3}]^T, A = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}, b = [4\ 3]^T$$

$$Aeq = [], Beq = [], lb = [0\ 0], ub = [].$$

- Solve it by Matlab: [x, fval]=quadprog(H,c,A,b,Aeq,Beq,lb,ub)

Quadratic Programming: standard form (2)

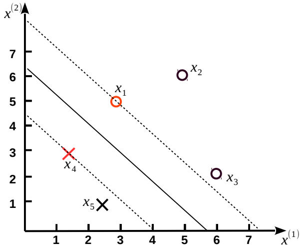
$$H = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}, c = [\text{-4 -3}]^T, A = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}, b = [4\ 3]^T$$

$$Aeq = [], Beq = [], lb = [0\ 0], ub = [].$$

- Solve it by Matlab: **quadprog**

1. H=[4 -1;-1 2];
2. c=[-4 -3]';
3. A=[1 1;-1 3];
4. b=[4 3]';
5. Aeq=[];Beq=[];
6. lb=[0 0]';ub=[]
7. [x fval]=quadprog(H,c,A,b,Aeq,Beq,lb,ub);

Simplified SVM: 2D case



$$\min_{w,b} \quad \frac{1}{2}||w||^2 \qquad (7)$$

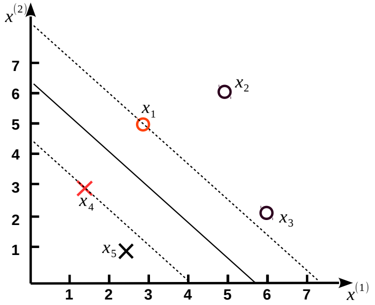
$$\text{s. t.} \quad 1 - y^{(i)}(w^T x^{(i)} + b) \leq 0$$

- In 2D case, $w = [w_1, w_2]$
- y=1 for positive example
- y=-1 for negative example
- Try to solve this problem with '**quadpgrog**'
- Display your result

- Positive: $x_1(3, 5)$; $x_2(5, 6)$; $x_3(6, 2.3)$
- Negative: $x_4(1.5, 3)$; $x_5(2.5, 1)$

- **Hints**: [x fval]=quadprog(H,c,A,B,Aeq,Beq,lb,ub);

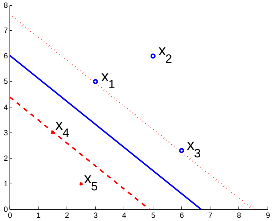
Simplified SVM: solution (1)



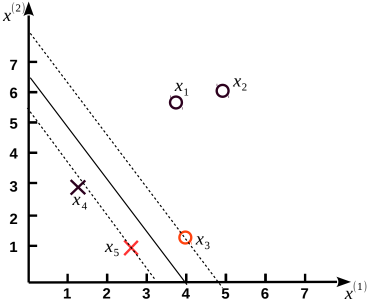
$$\min_{w,b} \quad \frac{1}{2}(w_1^2 + w_2^2) \tag{8}$$
$$\text{s.t.} \quad -3w_1 - 5w_2 - b \leq -1$$
$$-5w_1 - 6w_2 - b \leq -1$$
$$-6w_1 - 2.3w_2 - b \leq -1$$
$$1.5w_1 + 3w_2 + b \leq -1$$
$$2.5w_1 + w_2 + b \leq -1$$

- Positive: $x_1(3, 5)$; $x_2(5, 6)$; $x_3(6, 2.3)$
- Negative: $x_4(1.5, 3)$; $x_5(2.5, 1)$
- **Hints**: [x fval]=quadprog(H, c, A, B, Aeq, Beq, lb, ub);

Simplified SVM: result



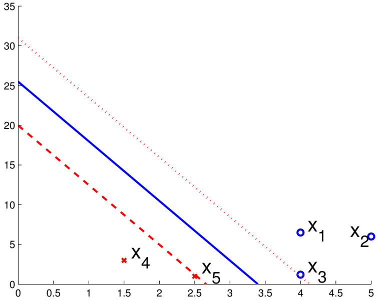
Change training points, see what happens



$$\begin{aligned} \underset{w,b}{\text{Min.}} \quad & \frac{1}{2}(w_1^2 + w_2^2) \qquad (9) \\ \text{s. t.} \quad & -4w_1 - 6.5w_2 - b \le -1 \\ & -5w_1 - 6w_2 - b \le -1 \\ & -4w_1 - 1.2w_2 - b \le -1 \\ & 1.5w_1 + 3w_2 + b \le -1 \\ & 2.5w_1 + w_2 + b \le -1 \end{aligned}$$

- Positive: $x_1(4, 6.5)$; $x_2(5, 6)$; $x_3(4, 1.2)$
- Negative: $x_4(1.5, 3)$; $x_5(2.5, 1)$
- **Hints**: [x fval]=quadprog(H, c, A, B, Aeq, Beq, lb, ub);

Simplified SVM: result updated



- This time x_3 and x_5 become the support vectors

Food for thought

1. Are all the training points used to determine the splitting line?
2. Is it possible to train the model with only negative or positive points?
3. The more training points the better?
4. Why in practice we do not solve SVM by QP?