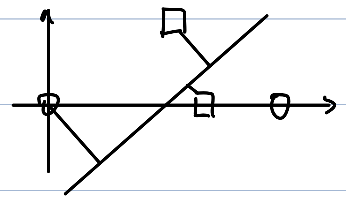
Lec 27

Recap: SVM

↳ $\min_{\underline{w}, b} \frac{1}{2} \underline{w}^T \underline{w}$ s.t. $y_n(\underline{w}^T \underline{x}_n + b) \geq 1$, for $n = 1, 2, \cdots N$



↳ constrained optimization problem

No analytical solution → Numerical

Objective : $\frac{1}{2} \sum_{i=1}^{d} w_i^2$ quarditic function in $\{w_i\}$

Constraints linear in $\{w_i\}$ and b

case of QP problems

1. Quaradic programming (QP) problems:

↳ Canonical form:

$$\min_{\underline{u} \in \mathbb{R}^L} \frac{1}{2} \underline{u}^T Q \underline{u} + \underline{p}^T \underline{u}$$

s.t. $\underline{a}_m^T \underline{u} \geq c_m$, for $m = 1, 2, \cdots M$

where $Q \in \mathbb{R}^{L \times L}$, $\underline{p} \in \mathbb{R}^L$, $\underline{a}_m \in \mathbb{R}^L$, $\forall m$ are pre-specified parameters.

↳ Note:

① $\underline{a}_m^T \underline{u} \geq c_m$ is general

e.g. $\underline{a}^T \underline{u} \leq 1 \iff -\underline{a}^T \underline{u} \geq -1$

$\underline{a}^T \underline{u} = 1 \iff \begin{cases} \underline{a}^T \underline{u} \geq 1 \\ -\underline{a}^T \underline{u} \geq -1 \end{cases}$ > two constraints

$u_i \geq 2 \iff [0, \cdots, 0, 1, 0, \cdots 0]\, \underline{u} \geq 2$

$\uparrow$ i th element

② The M constraints are often compactly written:

$$A\underline{u} \geq \underline{c}$$

where $A = \begin{bmatrix} \underline{a}_1^T \\ \vdots \\ \underline{a}_M^T \end{bmatrix}$, $\underline{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_M \end{bmatrix}$

③ If Q is positive semi-definite,

// i.e. $\underline{u}^T Q \underline{u} \geq 0$ for any $\underline{u} \in \mathbb{R}^L$

then the QP problem is convex $\Rightarrow$ easy to solve

otherwise hard to solve

④ Many exisiting software solvers

2. SVM is a convex QP problems.

$\hookrightarrow$ Let $\underline{u} = \begin{bmatrix} b \\ \underline{w} \end{bmatrix} \in \mathbb{R}^{d+1}$ $\quad (L = d+1)$

$$Q = \left[\begin{array}{c:cccc} 0 & 0 & \cdots & & 0 \\ \hdashline 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \cdots & 0 \\ & \vdots & & & \\ 0 & 0 & \cdots & & 1 \end{array}\right] = \left[\begin{array}{c:c} 0 & \underline{0}_d^T \\ \hdashline \underline{0}_d & I_d \end{array}\right]$$

$\because \underline{p} = \underline{0}_{d+1}$

$\therefore \frac{1}{2}\underline{u}^T Q \underline{u} + \underline{p}^T \underline{u} = \frac{1}{2}\underline{w}^T \underline{w}$

Let $\underline{a}_n^T = [y_n \cdots y_n \underline{x}_n^T]. \in \mathbb{R}^{d+1}$, $\quad c_n = 1$

then $\underline{a}_n^T \underline{u} \geq c_n \iff y_n b + y_n \underline{x}_n^T \underline{w} \geq 1$, $\quad \forall\, n = 1, 2, \cdots N$

$(M = N)$ $\qquad\qquad (\checkmark)$

↳ Convexity:

$$\underline{u}^T Q \underline{u} = \begin{bmatrix} \quad \end{bmatrix} \begin{bmatrix} \; & \; \\ \; & I \end{bmatrix} \begin{bmatrix} \; \end{bmatrix} = u_2^2 + u_3^2 + \cdots + u_{d+1}^2 \geq 0$$

$$\forall \; \underline{u} \in \mathbb{R}^L$$

To solve SVM problem, plug Q, $\underline{p}$, $\{\underline{a}_n\}$, $\{c_n\}$ above into any QP solver $\longrightarrow (\underline{w}^*, b^*)$

3. Optimal Soft-margin SVM:

(reduce over-fitting)

↳ Allow some margin violation $\xi_n \geq 0$, but penalize it to discourge large violation

$$\min_{\underline{w}, b, \underline{\xi}} \; \frac{1}{2} \underline{w}^T \underline{w} + C \sum_{n=1}^{N} \xi_n$$

coeff that how much you care the violation

$$\text{s.t. } y_n (\underline{w}^T x_n + b) \geq 1 - \xi_n, \quad n = 1, \cdots N$$

margin violation

↳ Note:

① Large C $\equiv$ we try to avoid penalty

⇒ close to hard-margin SVM

smaller margin & less violation

② still a convex QP problem

$$\underline{u} = \begin{bmatrix} b \\ \underline{w} \\ \underline{\xi} \end{bmatrix}, \quad Q = \begin{bmatrix} 0 & \underline{0}_d^T & \underline{0}_N^T \\ \underline{0}_d & I_d & 0_{d \times N} \\ \underline{0}_N & 0_{N \times d} & 0_{N \times N} \end{bmatrix}, \quad \underline{p} = \begin{bmatrix} \underline{0}_{d+1} \\ C \, \underline{1}_N \end{bmatrix}, \quad \underline{a}_m = \cdots, \quad c_m = \cdots$$