MLHW 5

Supporting Vector Regression (regression problem)

Given ① training set $\{(x_1, y_1), \cdots (x_m, y_m)\}$ where $x_i \in \mathbb{R}^{n+1}$, $y_i \in \mathbb{R}$

② $f(x) = w^T x + b \Rightarrow$ the function of the hyperplane.

③ convex optimization problem is

$$\underset{w,b,\xi}{\text{minimize}} \quad \frac{1}{2}\|w\|^2 + \underbrace{C}_{\text{penalty}} \sum_{i=1}^{m} \underbrace{\xi_i}_{\text{the distance of the hyperplane to all data points}}$$

such that

$$y_i - w^T x_i - b \le \epsilon + \xi_i$$
$$w^T x_i + b - y_i \le \epsilon + \xi_i \qquad \forall i = 1 \dots m$$
$$\xi_i \ge 0$$

④ given $\epsilon > 0$ fixed value, $C > 0$

$\boxed{Q1}$ Objective: write down the lagranian for the "optimization" problem with lagranian multiplier $\alpha_i, \alpha_i^*, \beta_i \Rightarrow \mathcal{L}(w, b, \xi, \alpha, \alpha^*, \beta)$

<sol> Recall to SVM's lagranian optimization problem, we have to rewrite the problem to dual and Primal form.

1. Rewrite SVR problem to primal form.

Primal Problem (P)

minimize $f(x)$

subject to $g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix} \le 0$, $h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_\ell(x) \end{bmatrix} = 0$

variable $x \in \mathcal{X}$

compare

$$\text{minimize} \quad f(w,b,\xi) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{\overset{m \text{ data size}}{m}} \xi_i$$

subject to

$$g_{1i}(w,b,\xi) = y_i - w^T x_i - b - \epsilon - \xi_i \le 0$$
$$g_{2i}(w,b,\xi) = w^T x_i + b - y_i - \epsilon - \xi_i \le 0$$
$$g_{3i}(w,b,\xi) = -\xi_i \le 0$$

Variables $(w, b, \xi) \in \mathbb{R}^{n+1} \times \mathbb{R} \times \mathbb{R}^m$

2. Analyze it and consider if it suit Strong Duality Constraint

(1) variables (w, b, ξ) is absolutly non-empty and convex in $\mathbb{R}^{n+1} \times \mathbb{R} \times \mathbb{R}^m$

(2) basically, $f = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i$ is a $\underbrace{\text{quadratic function}}_{\Downarrow \text{ convex}}$ + $\underbrace{\text{linear function}}_{\Downarrow \text{ convex}}$ = convex function

(Note that, convex function + convex function equal to convex function can prove in lecture!)

(3) g_{1i}, g_{2i}, g_{3i} are all linear $\rightarrow$ convex

(4) and (5) are related to equality constraint which is not included in the problem.

(6) Suppose there exists $\hat{x} \in \mathcal{X}$ such that $g(\hat{x}) < 0$ and $h(\hat{x}) = 0$ (slator's condition)

$\Rightarrow$ To fit this constraint, we must find a (w, b, ξ) such that inequality constraints are satisfied.

⇒ First, fixed w and b with random, then choose larger enough (e.g. ∞) to satisfied the all three constraint.

⇒ Therefore, we can indeed find a variable set to satisfied the constraints.

⇒ SVR can satisfy strong duality theorem.

3. Rewrite SVR problem to dual form.

Dual Problem (D)

maximize $\theta(u,v) = \inf\left[f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{l} v_j h_j(x) \right]$

subject to $u \geq 0$

variable $u = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} \in \mathbb{R}^m$, $v = \begin{bmatrix} v_1 \\ \vdots \\ v_l \end{bmatrix} \in \mathbb{R}^l$

compare

maximize $\theta(\alpha, \alpha^*, \beta) = \inf_{w \in \mathbb{R}^n, b \in \mathbb{R}, \xi \in \mathbb{R}^m} \left[f(w,b,\xi) + \sum_{i=1}^{m} \alpha_i g_{1i}(w,b,\xi) + \sum_{i=1}^{m} \alpha_i^* g_{2i}(w,b,\xi) + \sum_{i=1}^{m} \beta_i g_{3i}(w,b,\xi) \right]$

subject to $\alpha_1 \dots \alpha_m \geq 0$ (for 1st constraint)
$\alpha_1^* \dots \alpha_m^* \geq 0$ (for 2nd constraint)
$\beta_1 \dots \beta_m \geq 0$ (for 3rd constraint)

variable $(\alpha, \alpha^*, \beta) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$

⇒ $L(w,b,\xi,\alpha,\alpha^*,\beta) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i\left[y_i - w^T x_i - b - \epsilon - \xi_i \right] + \sum_{i=1}^{m} \alpha_i^*\left[w^T x_i + b - y_i - \epsilon - \xi_i \right] + \sum_{i=1}^{m} \beta_i \left[-\xi_i \right]$

⇒ $\theta(\alpha,\alpha^*,\beta) = \inf_{(w,b,\xi) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m} \left(L(w,b,\xi,\alpha,\alpha^*,\beta) \right)$

Q2.

4. Derivative

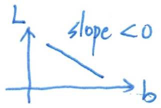
$\nabla_w L = w + \sum_{i=1}^{m} \alpha_i(-x_i) + \sum_{i=1}^{m} \alpha_i^* x_i$.

$\frac{\partial L}{\partial b} = \sum_{i=1}^{m} \alpha_i(-1) + \sum_{i=1}^{m} \alpha_i^*$

$\frac{\partial L}{\partial \xi_i} = C + \alpha_i(-1) + \alpha_i^*(-1) + \beta_i(-1)$

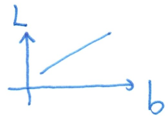
5. Analyze

- If $\sum_{i=1}^{m} \alpha_i^* - \sum_{i=1}^{m} \alpha_i < 0$, that is while b is increasing, L will decreasing



$$\Rightarrow \inf L(w, b, \xi, \alpha, \alpha^*, \beta) = -\infty$$

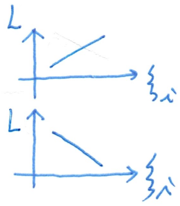
- If $\sum_{i=1}^{m} \alpha_i^* - \sum_{i=1}^{m} \alpha_i > 0$, that is while b is decreasing, L will decreasing



$$\Rightarrow \inf L(w, b, \xi, \alpha, \alpha^*, \beta) = -\infty$$

For some $i \Rightarrow$

- If $C - \alpha_i - \alpha_i^* - \beta_i > 0$, < 0, that is while ξ_i is decreasing / increasing, L is decreasing



$$\Rightarrow \inf L(w, b, \xi, \alpha, \alpha^*, \beta) = -\infty$$

$\Rightarrow$ We can conclude that "If $\sum_{i=1}^{m} \alpha_i^* - \sum_{i=1}^{m} \alpha_i \neq 0$ or $C \neq (\alpha_i + \alpha_i^* + \beta_i)$, then $\theta(\alpha, \alpha^*, \beta) = -\infty$"

6. Simplify

According to the observation above, the minimum value at $\nabla_w L = 0$ and $\sum_{i=1}^{m} \alpha_i^* - \sum_{i=1}^{m} \alpha_i = 0$ and $C = \alpha_i + \alpha_i^* + \beta_i$

$$\Rightarrow w = \sum_{i=1}^{m} \alpha_i x_i + \sum_{i=1}^{m} \alpha_i^* x_i$$

Substitude to $L(w, b, \xi, \alpha, \alpha^*, \beta)$

$$\Rightarrow \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i\left[y_i - w^T x_i - b - \epsilon - \xi_i\right] + \sum_{i=1}^{m} \alpha_i^*\left[w^T x_i + b - y_i - \epsilon - \xi_i\right] + \sum_{i=1}^{m} \beta_i(-\xi_i)$$

$$= \frac{1}{2}\left(\sum_{i=1}^{m} (\alpha_i + \alpha_i^*) x_i\right)^T \left(\sum_{j=1}^{m} (\alpha_j + \alpha_j^*) x_j\right) +$$

$$\sum_{i=1}^{m} \cancel{\alpha_i \xi_i} + \cancel{\alpha_i^* \xi_i} + \cancel{\beta_i \xi_i} + \cancel{\alpha_i y_i} - \alpha_i \cancel{w^T} x_i - \cancel{\alpha_i b} - \alpha_i \epsilon - \cancel{\alpha_i \xi_i} + \alpha_i^* \cancel{w^T} x_i + \cancel{\alpha_i^* b} - \cancel{\alpha_i^* y_i}$$

$$- \alpha_i^* \epsilon - \cancel{\alpha_i^* \xi_i} - \cancel{\beta_i \xi_i}$$

$w = \sum_{i=1}^{m} (\alpha_i + \alpha_i^*) x_i$

and $\sum_{i=1}^{m} \alpha_i^* - \alpha_i = 0$

and $C = \alpha_i + \alpha_i^* + \beta_i$

$$\cancel{= \frac{1}{2}\left(\sum_{i=1}^{m} (\alpha_i + \alpha_i^*) x_i\right)^T \left(\sum_{i=1}^{m} (\alpha_i + \alpha_i^*) x_i\right) + \sum_{i=1}^{m} \alpha_i (y_i - w^T x_i - b) - \alpha_i^* (y_i - w^T x_i - b) - \epsilon(\alpha_i + \alpha_i^*)}$$

$$\cancel{= \frac{1}{2} \sum_{i=1}^{m} (x_i^T \alpha_i^T + x_i^T \alpha_i^{*T})(\alpha_i x_i + \alpha_i^* x_i) - \epsilon(\alpha_i + \alpha_i^*)}$$

$$= \frac{1}{2} \sum_{i=1}^{m} (x_i^T \alpha_i^T + x_i^T \alpha_i^{*T}) \sum_{j=1}^{m} (\alpha_j x_j + \alpha_j^* x_j) - \sum_{i=1}^{m} \epsilon(\alpha_i + \alpha_i^*) \Rightarrow \text{cont.}$$

cont.

$$= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} x_i^T \alpha_i^T \alpha_j x_j + x_i^T \alpha_i^{*T} \alpha_j x_j + x_i^T \alpha_i^T \alpha_j^* x_j + x_i^T \alpha_i^{*T} \alpha_j^* x_j - \sum_{i=1}^{m} \epsilon(\alpha_i + \alpha_i^*)$$

maximize $\theta(\alpha, \alpha^*, \beta) =$

subject to $\alpha \geq 0$, $\alpha^* \geq 0$, $\beta \geq 0$

$$\sum_{i=1}^{m} \alpha_i^* - \alpha_i = 0$$

$$C = \alpha_i + \alpha_i^* + \beta_i$$

variable $(\alpha, \alpha^*, \beta) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$

simplify : subject to $\sum_{i=1}^{m} \alpha_i^* - \alpha_i = 0$, $\alpha \geq 0$, $\alpha^* \geq 0$

variables

$$0 \leq \alpha_i + \alpha_i^* \leq C$$

Q3 Given primal optimal solution $= (\bar{w}, \bar{b}, \bar{\xi})$, $\bar{w} = \sum_{i=1}^{m} (\bar{\alpha}_i - \bar{\alpha}_i^*) x_i$

dual optimal solution $= (\bar{\alpha}, \bar{\alpha}^*, \bar{\beta})$

(1) Objective : Prove $\bar{b} = \underset{b \in \mathbb{R}}{\operatorname{argmin}}\ C\sum_{i=1}^{m} \max\left[|y_i - (\bar{w}^T x_i + b)| - \epsilon, 0\right]$

The original function for SVR is :

$$f(w, b) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \max\left[|y_i - (w^T x_i + b)| - \epsilon, 0\right]$$

Then if we minimize $f(w,b)$ with given optimal w, that is we want to find optimal "b".

It's equally $\bar{b} = C\sum_{i=1}^{m} \max\left[|y_i - (\bar{w}^T x_i + b)| - \epsilon, 0\right]$

(2) Given $e = y_i - (\bar{w}^T x_i + \bar{b})$, Objective : prove vary situation

<sol> First, find KKT condition :

- Stationary Condition : (S1) $\nabla_w L = w - \sum_{i=1}^{m} (\alpha_i + \alpha_i^*) x_i = 0$,

  (S2) $\sum_{i=1}^{m} \alpha_i^* - \alpha_i = 0$, $\forall i = 1 \ldots m$.

  (S3) $C = \alpha_i + \alpha_i^* + \beta_i$,

- Primal & Dual feasibility Condition : (P1) $y_i - w^T x_i - b - \epsilon - \xi_i \leq 0$,

  (P2) $w^T x_i + b - y_i - \epsilon - \xi_i \leq 0$,

  (P3) $-\xi_i \leq 0$, $\forall i$

  (D1) $\alpha \geq 0$,

  (D2) $\alpha^* \geq 0$,

  (D3) $\beta \geq 0$,

- Complementary Slackness (C1) $\alpha_i (y_i - w^T x_i - b - \epsilon - \xi_i) \overset{(C2)}{=} \alpha_i^* (w^T x_i + b - y_i - \epsilon - \xi_i) \overset{(C3)}{=} \beta_i(-\xi_i) = 0, \forall i$

Second, analyze the situation.

- If $|y_i - (\bar{w}^T x_i + \bar{b})| < \epsilon$, then

$$\text{C1} \Rightarrow \alpha_i(\underbrace{\underbrace{\underbrace{e}_{<\epsilon} - \epsilon}_{<0} - \underbrace{\xi_i}_{\geq 0\ \text{(D3)}}}_{<0}) = 0 \Rightarrow \alpha_i \times \overset{<0}{\bigcirc} = 0 \overset{\text{imply}}{\Rightarrow} \boxed{\alpha_i = 0}$$

$$\text{C2} \Rightarrow \alpha_i^*(\underbrace{\underbrace{\underbrace{-e}_{<\epsilon} - \epsilon}_{<0} - \underbrace{\xi_i}_{\geq 0\ \text{(D3)}}}_{<0}) = 0 \Rightarrow \alpha_i^* \times \overset{<0}{\bigcirc} = 0 \overset{\text{imply}}{\Rightarrow} \boxed{\alpha_i^* = 0}$$

$$\text{S3} \Rightarrow C = \beta_i > 0 \text{ (initial constraint)} \Rightarrow \text{C3}\ \underbrace{\beta_i}_{>0}(-\xi_i) = 0 \overset{\text{imply}}{\Rightarrow} \boxed{\xi_i = 0}$$

- If $e = \epsilon$

$$\text{C1} \Rightarrow \alpha_i(e - \epsilon - \xi_i) = 0 \Rightarrow \alpha_i(-\xi_i) = 0 \overset{\text{C3}}{=} \beta_i(-\xi_i) \Rightarrow \alpha_i = \beta_i$$

$$\text{C2} \Rightarrow \alpha_i^*(\underbrace{\underbrace{\underbrace{-e-\epsilon}_{\substack{-2\epsilon \\ \downarrow \\ >0 \Rightarrow \text{initial constraint}}}}_{<0} - \underbrace{\xi_i}_{\substack{\downarrow \\ \geq 0}}}_{<0}) = 0 \Rightarrow \text{imply}\ \alpha_i^* = 0$$

$$\because \alpha_i = \beta_i \text{ and } \alpha_i^* = 0 \Rightarrow \text{S3} \Rightarrow C = \alpha_i + \beta_i \Rightarrow C(-\xi_i) = \underbrace{(\alpha_i + \beta_i)}_{>0}(-\xi_i) = 0 \overset{\text{imply}}{\Rightarrow} \boxed{\xi_i = 0}$$

$$\because \text{D1}, \text{D3}, \text{S3} \begin{cases} \alpha \geq 0 \\ \beta \geq 0 \\ C = \alpha_i + \alpha_i^* + \beta_i \end{cases} \Rightarrow \therefore \boxed{0 \leq \alpha_i \leq C.}$$

- If $e = -\epsilon$

$$\text{C1} \Rightarrow \alpha_i(\underbrace{\underbrace{\underbrace{e-\epsilon}_{-2\epsilon}}_{<0} - \underbrace{\xi_i}_{\geq 0\ \text{(D3)}}}_{<0}) = 0 \overset{\text{imply}}{\Rightarrow} \alpha_i = 0$$

$$\text{C2} \Rightarrow \alpha_i^*(-e - \epsilon - \xi_i) = 0 \Rightarrow \alpha_i^*(-\xi_i) = 0 \overset{\text{C3}}{=} \beta_i(-\xi_i)$$

$$\text{S3} \Rightarrow C = \underbrace{\alpha_i}_{0} + \alpha_i^* + \beta_i = \alpha_i^* + \beta_i$$

$$\Rightarrow C(-\xi_i) = \underbrace{(\alpha_i^* + \beta_i)}_{>0}(-\xi_i) = 0 \Rightarrow \boxed{\xi_i = 0}$$

$$\because \text{D2}, \text{D3}, \text{S3} \Rightarrow \begin{cases} \alpha^* \geq 0 \\ \beta \geq 0 \\ C = \underbrace{\alpha_i}_{0} + \alpha_i^* + \beta_i \end{cases} \Rightarrow \therefore \boxed{0 \leq \alpha_i^* \leq C}$$

- If $e > \epsilon$

$$C_2 \Rightarrow \alpha_i^*(\underbrace{\underbrace{-e-\epsilon}_{<0} - \underbrace{\xi_i}_{\geq 0}}_{<0}) = 0 \overset{\text{imply}}{\Rightarrow} \alpha_i^* = 0$$

$$P_1 \Rightarrow \underbrace{e-\epsilon}_{>0} - \xi_i \leq 0 \overset{\text{imply}}{\Rightarrow} \xi_i > 0 \overset{C_3}{\Rightarrow} \beta_i(-\underbrace{\xi_i}_{>0}) = 0 \overset{\text{imply}}{\Rightarrow} \beta_i = 0$$

$$S_3 \Rightarrow C = \alpha_i + \underbrace{\alpha_i^*}_{=0} + \underbrace{\beta_i}_{=0} \overset{\text{imply}}{\Rightarrow} \boxed{C = \alpha_i}$$

$$C_1 \Rightarrow \underbrace{\alpha_i}_{>0}(e-\epsilon-\xi_i) = 0 \overset{\text{imply}}{\Rightarrow} e-\epsilon-\xi_i = 0 \Rightarrow \boxed{\xi_i = e-\epsilon}$$

- If $e < -\epsilon$

$$C_1 \Rightarrow \alpha_i(\underbrace{\underbrace{e-\epsilon}_{<0} - \underbrace{\xi_i}_{\geq 0}}_{<0}) = 0 \overset{\text{imply}}{\Rightarrow} \alpha_i = 0$$

$$P_2 \Rightarrow \underbrace{-e-\epsilon}_{>0} - \xi_i \leq 0 \overset{\text{imply}}{\Rightarrow} \xi_i > 0 \overset{C_3}{\Rightarrow} \beta_i(-\xi_i) = 0 \overset{\text{imply}}{\Rightarrow} \beta_i = 0$$

$$S_3 \Rightarrow C = \underbrace{\alpha_i}_{=0} + \alpha_i^* + \underbrace{\beta_i}_{=0} \overset{\text{imply}}{\Rightarrow} \boxed{C = \alpha_i^*}$$

$$C_2 \Rightarrow \underbrace{\alpha_i^*}_{>0}(-e-\epsilon-\xi_i) = 0 \overset{\text{imply}}{\Rightarrow} -e-\epsilon-\xi_i = 0 \Rightarrow \boxed{\xi_i = -(e+\epsilon)}$$

Q.E.D.