Support Vector Regression

S. Sumitra
Department of Mathematics
Indian Institute of Space Science and Technology

Advanced Machine Learning

Sparse Solution

- SVM classification is a sparse algorithm. The final solution consists of a subset of training points called support vectors support vectors.
- Such a representation has computational advantageous
- SVM regression use ϵ insenstive loss function to ensure sparseness of the solution

Error Tolerance: ϵ

- No error
  - $|y - \tilde{f}(x)| \leq \epsilon$
  - $-\epsilon \leq y - \tilde{f}(x) \leq \epsilon$
- Error
  - $|y - \tilde{f}(x))| > \epsilon$
  - $y - \tilde{f}(x) < -\epsilon$ or $y - \tilde{f}(x) > \epsilon$

ϵ insensitive loss function

$$L^{\epsilon}(x, y, \tilde{f}) = |y - \tilde{f}(x)|_{\epsilon} = \max(0, |y - \tilde{f}(x)| - \epsilon)$$

- $\max(0, (y - \tilde{f}(x) - \epsilon)) = \xi$, $\max(0, (\tilde{f}(x) - y) - \epsilon)) = \xi^*$
- ξ and ξ^* cannot be cannot be greater than zero at the same time.

Data

$$\{(x_1, y_1), (x_2, y_2) \ldots (x_N, y_N)\}, x_i \in \mathbb{R}^n, y_i \in \mathbb{R}$$

$$\min_{f \in \mathcal{F}, b \in \mathbb{R}} ||f||^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

subject to

$$(y_i - \tilde{f}(x_i)) - \epsilon \leq \xi_i, \ i = 1, \ldots, N$$

$$(\tilde{f}(x_i) - y_i) - \epsilon \leq \xi_i^*, \ i = 1, \ldots, N$$

$$\xi_i \geq 0, \ i = 1, \ldots, N$$

$$\xi_i^* \geq 0, \ i = 1, \ldots, N$$

Case 1: Linear

- $\tilde{f}(x) = \langle w, x \rangle + b, w \in \mathbb{R}^n, b \in \mathbb{R}$

$$\min_{w \in \mathbb{R}^n, b \in \mathbb{R}} ||w||^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

subject to

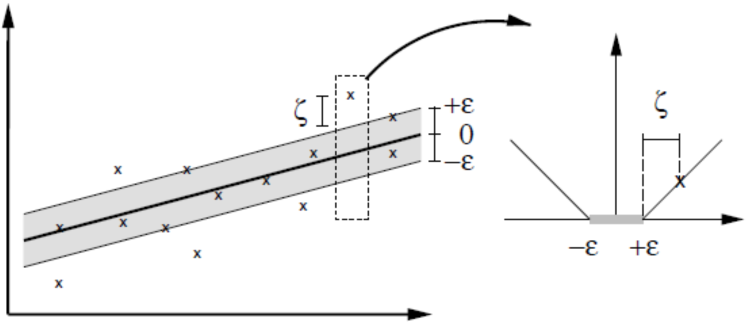
$$(y_i - \langle w, x_i \rangle - b) - \epsilon \leq \xi_i, \; i = 1, \ldots, N$$

$$(\langle w, x_i \rangle + b - y_i) - \epsilon \leq \xi_i^*, \; i = 1, \ldots, N$$

$$\xi_i \geq 0, \; i = 1, \ldots, N$$

$$\xi_i^* \geq 0, \; i = 1, \ldots, N$$

Case 1: Hyperplane



$$\min_{w \in \mathbb{R}^n, b \in \mathbb{R}} ||w||^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

subject to

$$\xi_i - y_i + \langle w, x_i \rangle + b + \epsilon \geq 0, \; i = 1, \ldots, N$$

$$\xi_i^* - \langle w, x_i \rangle - b + y_i + \epsilon \geq 0, \; i = 1, \ldots, N$$

$$\xi_i \geq 0, \; i = 1, \ldots, N$$

$$\xi_i^* \geq 0, \; i = 1, \ldots, N$$

Lagrangian Formulation

The Lagrangian can be written as

$$\begin{aligned} L &= \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\alpha_i(\epsilon + \xi_i - y_i + \langle w, x_i\rangle + b) \\ &\quad - \sum_{i=1}^{N}\alpha_i^*(\epsilon + \xi_i^* + y_i - \langle w, x_i\rangle - b) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*) \end{aligned}$$

For $i = 1, 2, \ldots N$,

$$\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$$

KKT Conditions

$$\partial_w L = w - \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) x_i = 0$$

$$w = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) x_i \tag{1}$$

$$\partial_b L = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) = 0 \tag{2}$$

$$\partial_{\xi_i} = C - \alpha_i - \eta_i = 0, i = 1, 2, \ldots N \tag{3}$$

$$\partial_{\xi_i^*} = C - \alpha_i^* - \eta_i^* = 0, , i = 1, 2, \ldots N \tag{4}$$

KKT Complimentary Conditions

For $i = 1, 2, \ldots N$

$$\alpha_i(\epsilon + \xi_i - y_i + \langle w, x_i \rangle + b) = 0 \quad (5)$$

$$\alpha_i^*(\epsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) = 0 \quad (6)$$

$$\eta_i \xi_i = 0 \quad (7)$$

$$\eta_i^* \xi_i^* = 0 \quad (8)$$

From (3),

$$0 \leq \alpha_i \leq C \tag{9}$$

From (13),

$$0 \leq \alpha_i^* \leq C \tag{10}$$

α_i and α_i^* cannot be greater than zero at the same time.

When $\alpha_i = 0$, from (3), $\eta_i > 0$. Therefore from (7), $\xi_i = 0$
Therefore from (5),

$$\epsilon - y_i + \langle w, x_i \rangle + b \geq 0$$

That is

$$y_i - \langle w, x_i \rangle - b \leq \epsilon$$

$H : y_i - (\langle w, x \rangle + b) = 0$, $H_1 : y_i - (\langle w, x \rangle + b) = \epsilon$,
$H_2 : (\langle w, x \rangle + b) - y_i = \epsilon$
That is (x_i, y_i) lies on H or on H_1 or between H and H_1
That is, in this case, (x_i, y_i) lies on or inside ϵ- insensitive tube.

Similarly $\alpha_i^* = 0$, from (13), $\eta_i^* > 0$. Therefore from (17), $\xi_i^* = 0$
Therefore from (6),

$$\epsilon + y_i - \langle w, x_i \rangle - b \geq 0$$

That is

$$\langle w, x_i \rangle + b - y_i \leq \epsilon$$

That is (x_i, y_i) lies on H or on H_2 or between H and H_2
That is, in this case, (x_i, y_i) lies on or inside ϵ tube.

When $0 < \alpha_i < C$, from (3), $\eta_i > 0$. Therefore from (7), $\xi_i = 0$
Therefore from (5),

$$\epsilon - y_i + \langle w, x_i \rangle + b = 0$$

That is (x_i, y_i) lies on H_1 .

When $0 < \alpha_i^* < C$, from (13), $\eta_i^* > 0$. Therefore from (17), $\xi_i^* = 0$
Therefore from (6),

$$\langle w, x_i \rangle + b - y_i = \epsilon$$

That is (x_i, y_i) lies on H_2

When $\alpha_i = C$, from (3), $\eta_i = 0$. Therefore from (7), $\xi_i \geq 0$
Therefore from (5),

$$\epsilon - y_i + \langle w, x_i \rangle + b + \xi_i = 0$$

That is

$$y_i - \langle w, x_i \rangle - b = \epsilon + \xi_i$$

That is,

$$y_i - \langle w, x_i \rangle - b \geq \epsilon$$

That is (x_i, y_i) lies on H_1 or outside the ϵ tube

When $\alpha_i^* = C$, from (13), $\eta_i^* = 0$. Therefore from (17), $\xi_i^* \geq 0$
Therefore from (6),

$$\epsilon - \langle w, x_i \rangle - b + y_i + \xi_i = 0$$

That is

$$\langle w, x_i \rangle + b - y_i = \epsilon + \xi_i$$

That is,

$$\langle w, x_i \rangle + b - y_i \geq \epsilon$$

That is (x_i, y_i) lies on H_2 or outside the ϵ tube

Support Vectors

- Data with non zero coefficients in the expansion of w are called support vectors
- Support vectors lie on H_1 or on H_2 or outside the ϵ tube
- Coefficients vanish if the data points lie inside the ϵ tube

Coefficients: Primal Variable

- The terms that involve w : $\frac{1}{2}\|w\|^2 - \sum_i(\alpha_i - \alpha_i^*)\langle w, x_i\rangle$
- The terms that involve b : $b(\sum_i(\alpha_i^* - \alpha_i))$
- The terms that involve ξ_i : $\sum_{i=1}^N \xi_i(C - \alpha_i - \mu_i)$
- The terms that involve ξ_i^* : $\sum_{i=1}^N \xi_i(C - \alpha_i^* - \mu_i^*)$
- The terms that involve ϵ: $-\epsilon(\alpha_i + \alpha_i^*)$
- The remaining terms are: $\sum_i y_i(\alpha_i - \alpha_i^*)$

Coefficients: Primal Variable

$$
\begin{aligned}
\|w\|^2 &= \langle \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) x_i, \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) x_i \rangle \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle
\end{aligned}
$$

$$
\begin{aligned}
\sum_i (\alpha_i - \alpha_i^*) \langle w, x_i \rangle &= \sum_i (\alpha_i - \alpha_i^*) \left\langle \sum_j (\alpha_j - \alpha_j^*) x_j, x_i \right\rangle \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle
\end{aligned}
$$

Dual Formulation

maximize

$$\frac{-1}{2}\sum_{i,j}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle x_i, x_j\rangle - \epsilon\sum_i(\alpha_i+\alpha_i^*) + \sum_i y_i(\alpha_i-\alpha_i^*)$$

subject to $\sum_i(\alpha_i - \alpha_i^*) = 0$ and $\alpha_i, \alpha_i^* \in [0, C]$).

Determination of b

Determine b using the points for which

- $0 < \alpha_i < C$
- $0 < \alpha_i^* < C$

Case II: General Form

- $\tilde{f}(x) = \langle f, k_x \rangle + b$

$$\min_{f \in \mathcal{F}, b \in \mathbb{R}} ||f||^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

subject to

$$(y_i - \langle f, k_{x_i} \rangle - b) - \epsilon \leq \xi_i, \; i = 1, \ldots, N$$
$$(\langle f, k_{x_i} \rangle + b - y_i) - \epsilon \leq \xi_i^*, \; i = 1, \ldots, N$$
$$\xi_i \geq 0, \; i = 1, \ldots, N$$
$$\xi_i^* \geq 0, \; i = 1, \ldots, N$$

$$\min_{f \in \mathcal{H}, b \in \mathbb{R}} ||f||^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

subject to

$$\xi_i - y_i + \langle f, k_{x_i} \rangle + b + \epsilon \geq 0, \; i = 1, \ldots, N$$

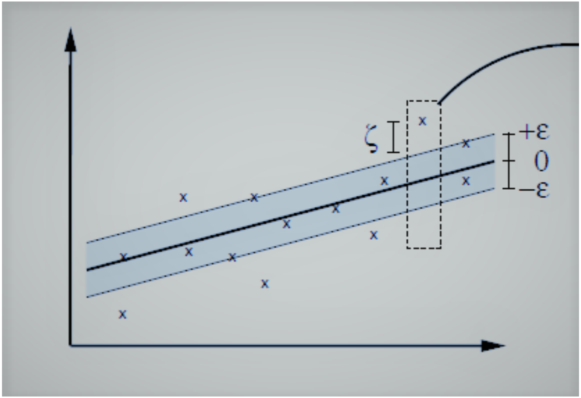
$$\xi_i^* - \langle f, k_{x_i} \rangle - b + y_i + \epsilon \geq 0, \; i = 1, \ldots, N$$

$$\xi_i \geq 0, \; i = 1, \ldots, N$$

$$\xi_i^* \geq 0, \; i = 1, \ldots, N$$

Hyperplane: RKHS

$H : y_i - (\langle f, k_{x_i} \rangle + b) = 0$, $H_1 : y_i - (\langle f, k_{x_i} \rangle + b) = \epsilon$,
$H_2 : (\langle f, k_{x_i} \rangle + b) - y_i = \epsilon$



Lagrangian Formulation

The Lagrangian can be written as

$$
\begin{aligned}
L &= \frac{1}{2}||f||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\alpha_i(\epsilon + \xi_i - y_i + \langle f, k_{x_i}\rangle + b) \\
&\quad - \sum_{i=1}^{N}\alpha_i^*(\epsilon + \xi_i^* + y_i - \langle f, k_{x_i}\rangle - b) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*)
\end{aligned}
$$

For $i = 1, 2, \ldots N$,

$$\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$$

KKT Conditions

$$\partial_f L = f - \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) k_{x_i} = 0$$

$$f = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) k_{x_i} \quad (11)$$

$$\partial_b L = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) = 0 \quad (12)$$

$$\partial_{\xi_i} = C - \alpha_i - \eta_i = 0, i = 1, 2, \ldots N \quad (13)$$

$$\partial_{\xi_i^*} = C - \alpha_i^* - \eta_i^* = 0, , i = 1, 2, \ldots N \quad (14)$$

KKT Complimentary Conditions

For $i = 1, 2, \ldots N$

$$\alpha_i(\epsilon + \xi_i - y_i + \langle f, k_{x_i} \rangle + b) = 0 \quad (15)$$

$$\alpha_i^*(\epsilon + \xi_i^* + y_i - \langle f, k_{x_i} \rangle - b) = 0 \quad (16)$$

$$\eta_i \xi_i = 0 \quad (17)$$

$$\eta_i^* \xi_i^* = 0 \quad (18)$$

$\alpha_i \alpha_i^* = 0$, there can never be a set of dual variables α_i, α_i^* which are both simultaneously nonzero.

From (13),

$$0 \leq \alpha_i \leq C \tag{19}$$

From (14),

$$0 \leq \alpha_i^* \leq C \tag{20}$$

α_i and α_i^* cannot be greater than zero at the same time.

When $\alpha_i = 0$, from (13), $\eta_i > 0$. Therefore from (17), $\xi_i = 0$
Therefore from (15),

$$\epsilon - y_i + \langle f, k_{x_i} \rangle + b \geq 0$$

That is

$$y_i - \langle f, k_{x_i} \rangle - b \leq \epsilon$$

That is (k_{x_i}, y_i) lies on H or on H_1 or between H and H_1
That is, in this case, (k_{x_i}, y_i) lies on or inside ϵ- insensitive tube.

Similarly when $\alpha_i^* = 0$, from (14), $\eta_i^* > 0$. Therefore from (18), $\xi_i^* = 0$
Therefore from (16),

$$\epsilon + y_i - \langle f, k_{x_i} \rangle - b \geq 0$$

That is

$$\langle f, k_{x_i} \rangle + b - y_i \leq \epsilon$$

That is (k_{x_i}, y_i) lies on H or on H_2 or between H and H_2
That is, in this case, (k_{x_i}, y_i) lies on or inside ϵ tube.

When $0 < \alpha_i < C$, from (13), $\eta_i > 0$. Therefore from (17), $\xi_i = 0$
Therefore from (15),

$$\epsilon - y_i + \langle f, k_{x_i} \rangle + b = 0$$

That is (k_{x_i}, y_i) lies on H_1.

When $0 < \alpha_i^* < C$, from (14), $\eta_i^* > 0$. Therefore from (18), $\xi_i^* = 0$.
Therefore from (16),

$$\langle f, k_{x_i} \rangle + b - y_i = \epsilon$$

That is (k_{x_i}, y_i) lies on H_2.

When $\alpha_i = C$, from (13), $\eta_i = 0$. Therefore from (17), $\xi_i \geq 0$ Therefore from (15),

$$\epsilon - y_i + \langle f, k_{x_i} \rangle + b + \xi_i = 0$$

That is

$$y_i - \langle f, k_{x_i} \rangle - b = \epsilon + \xi_i$$

That is,

$$y_i - \langle f, k_{x_i} \rangle - b \geq \epsilon$$

That is (k_{x_i}, y_i) lies on H_1 or outside the ϵ tube

When $\alpha_i^* = C$, from (14), $\eta_i^* = 0$. Therefore from (18), $\xi_i^* \geq 0$
Therefore from (16),

$$\epsilon - \langle f, k_{x_i} \rangle - b + y_i + \xi_i^* = 0$$

That is

$$\langle f, k_{x_i} \rangle + b - y_i = \epsilon + \xi_i^*$$

That is,

$$\langle f, k_{x_i} \rangle + b - y_i \geq \epsilon$$

That is (k_{x_i}, y_i) lies on H_2 or outside the ϵ tube

Support Vectors

- Data with non zero coefficients in the expansion of f are called support vectors
- Support vectors lie on H_1 or on H_2 or outside the ϵ tube
- Coefficients vanish if the data points lie inside the ϵ tube

Coefficients: Primal Variable

- The terms that involve f : $\frac{1}{2}\|f\|^2 - \sum_i(\alpha_i - \alpha_i^*)\langle f, k_{x_i}\rangle$
- The terms that involve b : $b(\sum_i(\alpha_i^* - \alpha_i))$
- The terms that involve ξ_i : $\sum_{i=1}^N \xi_i(C - \alpha_i - \mu_i)$
- The terms that involve ξ_i^* : $\sum_{i=1}^N \xi_i(C - \alpha_i^* - \mu_i^*)$
- The terms that involve ϵ: $-\epsilon(\alpha_i + \alpha_i^*)$
- The remaining terms are: $\sum_i y_i(\alpha_i - \alpha_i^*)$

Coefficients: Primal Variable

$$
\begin{aligned}
\|f\|^2 &= \langle \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) k_{x_i}, \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) k_{x_i} \rangle \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle k_{x_i}, k_{x_j} \rangle
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i} (\alpha_i - \alpha_i^*) \langle f, k_{x_i} \rangle &= \sum_{i} (\alpha_i - \alpha_i^*) \left\langle \sum_{j} (\alpha_j - \alpha_j^*) k_{x_j}, k_{x_i} \right\rangle \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle k_{x_i}, k_{x_j} \rangle
\end{aligned}
$$

Dual Formulation

maximize

$$\frac{-1}{2}\sum_{i,j}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle k_{x_i},k_{x_j}\rangle-\epsilon\sum_i(\alpha_i+\alpha_i^*)+\sum_i y_i(\alpha_i-\alpha_i^*)$$

subject to $\sum_i(\alpha_i-\alpha_i^*)=0$ and $\alpha_i,\alpha_i^*\in[0,C]$).

Determination of b

Determine b using the points for which

- $0 < \alpha_i < C$
- $0 < \alpha_i^* < C$