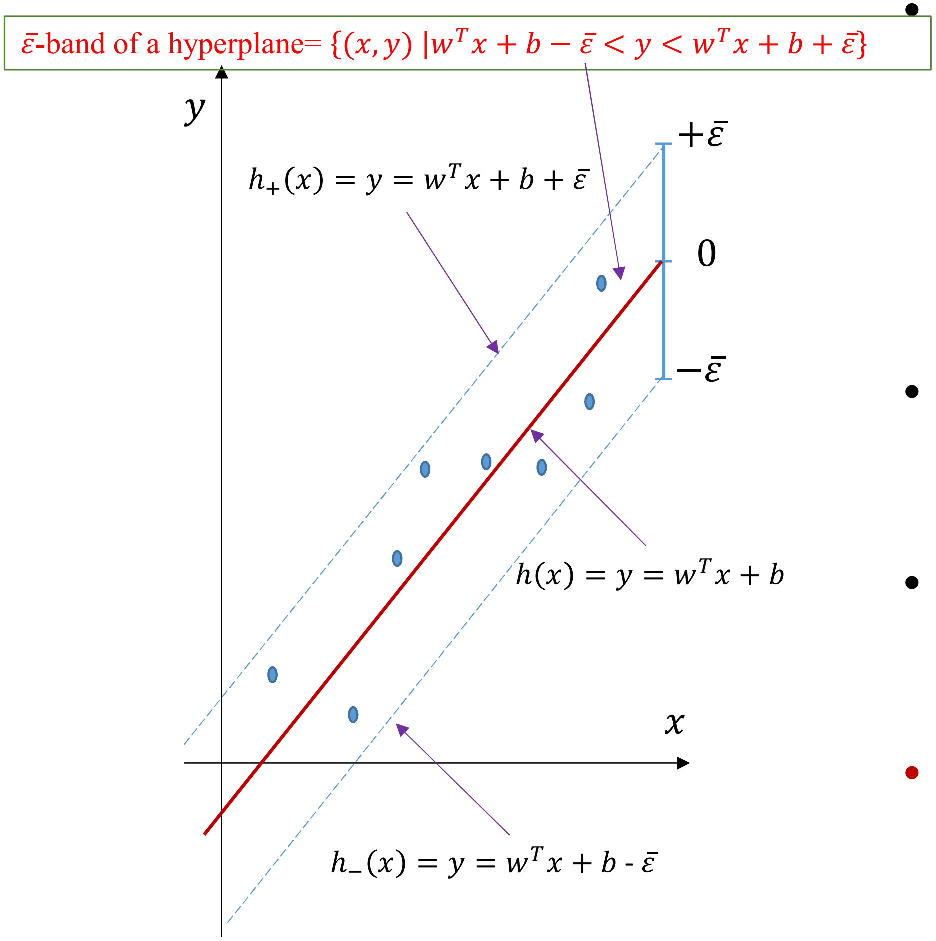
Support Vector machine
Linear Regression

Professor Abdellatif El Afia

Plan

- linear Regression Problems
- Hard $\bar{\varepsilon}$-Band Hyperplane
- Linear Hard ε-Band Support Vector Regression
- Linear ε-Band Support Vector Regression

linear Regression Problems



- Geometricaly, a linear regression problem in d-dimensional space corresponds to find a heperplane in (d+1)-dimensional space for given trainning set: $\{(x_i, y_i\}_{i=1}^{n}$ such that $\forall i :\ x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$

- Since a linear function defined in n-dimensional space is equivalent to hyperplane in the $\mathbb{R}^d \times \mathbb{R}$

- Our goal is to find a straight line with a small deviation from these points.

- Definition: we say that a hyperplane is the Hard $\bar{\varepsilon}$-band hyperplane for the training set S, if all the training points are in side its $\bar{\varepsilon}$-band

Hard $\bar{\varepsilon}$-band hyperplane

For a given $\bar{\varepsilon} > 0$ and a trainning set $\mathrm{S} = \{(x_i, y_i\}_{i=1}^{n}$,

Compute the optimal value ε_{inf} of the following problem

$$\bar{\varepsilon}\text{–band}_{inf} = \begin{cases} \qquad\qquad min \quad \bar{\varepsilon} \\ s.\text{à} \quad \bar{\varepsilon} \leq y_i - (w^T x_i - b) \leq \bar{\varepsilon}, i = 1, \dots, n \\ \qquad\qquad (w, b) \in \mathbb{R}^d \times \mathbb{R}\,, \bar{\varepsilon} > 0 \end{cases}$$

If $\bar{\varepsilon} > \varepsilon_{inf}$ then the Hard $\bar{\varepsilon}$-band hyperplane exist, and not uniquely

If $(\bar{\varepsilon} \leq \varepsilon_{inf})$ then there doesn't exist any Hard $\bar{\varepsilon}$-band hyperplane

Hard $\bar{\varepsilon}$-band hyperplane: constructing

For a given $\bar{\varepsilon} > \varepsilon_{inf} > 0$ and a training set $S = \{(x_i, y_i)\}_{i=1}^{n}$,

There exist a lot of hard $\bar{\varepsilon} - tube$ hyperplanes, However, which one is the best?

We construit tow classes $D^+ = \{(x_i, y_i + \bar{\varepsilon}), i = 1, \dots, n\}$ and $D^- = \{(x_i, y_i - \bar{\varepsilon}), i = 1, \dots, n\}$

then, the training set for classification is $S_{\bar{\varepsilon}} = \{(x_i, y_i + \bar{\varepsilon}, 1), (x_i, y_i - \bar{\varepsilon}, -1)\}_{i=1}^{n}$

Theorem: For a given $\bar{\varepsilon} > 0$ and a training set S, a hyperplane $y = w^T x + b$ is a Hard $\bar{\varepsilon}$-band hyperplane if and only the sets $D^+$and D^- locate on both sides of this hyperplane respectively and all of the points in $D^+$and D^- don't touch this hyperplane

Theorem shows that the better the Hard $\bar{\varepsilon}$-band hyperplane, the better the separating hyperplane with the training set S. So we can construct a Hard $\bar{\varepsilon}$-band hyperplane the using the classification method.

$$\text{Hard } \bar{\varepsilon}\text{–band hyperplane} \begin{cases} \min \quad \frac{1}{2} w^T w + \frac{1}{2} \eta^2 & \\ s.à \quad w^T x_i + \eta(y_i + \bar{\varepsilon}) + b \geq 1, & i = 1, \dots, n \\ \qquad w^T x_i + \eta(y_i - \bar{\varepsilon}) + b \leq -1, & i = 1, \dots, n \end{cases}$$

- $x_i \leftarrow (x_i, y_i + \bar{\varepsilon})$
- $w^T \leftarrow (w^T, \eta)$
- $w^T x_i \leftarrow (w^T, \eta)(x_i, y_i + \bar{\varepsilon}) = w^T x_i + \eta(y_i + \bar{\varepsilon})$
- $w^T x_i + b \leftarrow w^T x_i + \eta(y_i + \bar{\varepsilon}) + \mathrm{b}$
- $\frac{1}{2}\|w\|^2 \leftarrow \frac{1}{2}\|w\|^2 + \frac{1}{2}\eta^2$

Hard $\bar{\varepsilon}$-band hyperplane: constructing

Theorem:

Suppose that $(\bar{w}, \bar{b}, \bar{\eta})$ is the solution the problem Hard $\bar{\varepsilon}$-band hyperplane , then $\bar{\eta} \neq 0.$
Furthermore, let $\varepsilon = \bar{\varepsilon} - \frac{1}{\bar{\eta}}$ then:

- $\varepsilon_{inf} \leq \varepsilon < \bar{\varepsilon}$
- $(w^*, b^*) = \left(-\frac{\bar{w}}{\bar{\eta}}, -\frac{\bar{b}}{\bar{\eta}}\right)$ is the solution to the fallowing problem

$$\text{Hard } \varepsilon\text{–band hyperplane} \begin{cases} \min \quad \frac{1}{2} w^T w \\ y_i - (w^T x_i + b) \leq \varepsilon, \qquad i = 1, \ldots, n \\ s.à \quad w^T x_i + b - y_i \quad \leq \varepsilon, \qquad i = 1, \ldots, n \\ (w, b) \in \mathbb{R}^d \times \mathbb{R} \end{cases}$$

Linear Hard ε-band Support Vector Regression(SVR)

- Primal Problem

$$\mathrm{LH}-\varepsilon\text{–band}-SVR:\begin{cases} \quad min \quad \dfrac{1}{2}w^T w \\ \qquad y_i-(w^T x_i-b)\le\varepsilon, \qquad i=1,\dots,n \\ s.à \quad w^T x_i+b-y_i \quad \le\varepsilon, \qquad i=1,\dots,n \\ \qquad (w,b)\in\mathbb{R}^d\times\mathbb{R} \end{cases}$$

Theorem:

Suppose that ε_{inf} is the optimal value of the $\bar{\varepsilon}\text{–band}_{inf}$ problem

If $\varepsilon>\varepsilon_{inf}$, then the $\mathrm{LH}-\varepsilon\text{–band}-SVR$ problem has solutions, and the solution with respect to w is unique

Remark It's not necessarily true that the solution to Primal Problem with respect to b. In fact, when ε is large enough, there exist many b^* with different values, such that $(w^*,b^*)=(0,b^*)$ are the solutions.

Linear Hard $\boldsymbol{\varepsilon}$-band SVR : Primal Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$
- Choose an apprjudge parameter $\varepsilon > 0$
- Contruct and solve the optimzation problem $\mathrm{LH} - \varepsilon\text{–band} - SVR$ obtaining (w^*, b^*)

$$\mathrm{LH} - \varepsilon\text{–band} - SVR: \begin{cases} \quad min \quad \dfrac{1}{2} w^T w \\ \qquad y_i - (w^T x_i + b) \leq \varepsilon, \qquad i = 1, \dots, n \\ s.\text{à} \quad w^T x_i + b - y_i \quad \leq \varepsilon, \qquad i = 1, \dots, n \\ \qquad (w, b) \in \mathbb{R}^d \times \mathbb{R} \end{cases}$$

- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = (w^*)^T x + b^*$$

Linear Hard $\bar{\varepsilon}$-band SVR: Dual Form

- In order to drive the dual Form, we introduce the Lagrange Function

$$L_{LHSVR}(w,b,\lambda,\mu) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} \lambda_i(w^T x_i + b - y_i - \varepsilon) + \sum_{i=1}^{n} \mu_i(y_i - w^T x_i - b - \varepsilon)$$

According to chapter 1, the dual problem should have a form of

$$DLH - \varepsilon\text{–band} - SVR \begin{cases} Max \quad g(\lambda,\mu) = \inf\limits_{w\in\mathbb{R}^d, b\in\mathbb{R}} L_{LHSVR}(w,b,\lambda,\mu\,) \\ s.t \qquad \lambda_i \geq 0, \mu_i \geq 0, \qquad i = 1,..,n \end{cases}$$

As $L_{LHSVR}(w,b,\lambda,\mu)$is strictly convex quadratic Function of (w,b), its minimal value is achieved at (w,b)satisfying

- $\nabla_w L_{LHSVR}(w,b,\lambda,\mu) = w + \sum_{i=1}^{n}(\lambda_i - \mu_i)\, x_i = 0 \Longrightarrow w = -\sum_{i=1}^{n}(\lambda_i - \mu_i)\, x_i$
- $\nabla_b L_{LHSVR}(w,b,\lambda,\mu) = \sum_{i=1}^{n}(\lambda_i - \mu_i) = 0$

Linear Hard $\bar{\varepsilon}$-band SVR: Dual Form

One has by substitution in $L_{LHSVR}(w, b, \lambda, \mu)$:

$$L_{LHSVR}(w, b, \lambda, \mu) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\lambda_j - \mu_j)(\lambda_i - \mu_i)\left(x_j{}^T x_i\right) + \sum_{i=1}^{n}(\lambda_i - \mu_i)y_i - b\sum_{i=1}^{n}(\lambda_i - \mu_i) - \varepsilon\sum_{i=1}^{n}(\lambda_i + \mu_i)$$

If $\sum_{i=1}^{n}(\lambda_i - \mu_i) = 0$ then

$$\inf_{w\in\mathbb{R}^d, b\in\mathbb{R}} L_{LHSVR}(w, b, \lambda, \mu) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\lambda_j - \mu_j)(\lambda_i - \mu_i)\left(x_j{}^T x_i\right) + \sum_{i=1}^{n}(\lambda_i - \mu_i)y_i - \varepsilon\sum_{i=1}^{n}(\lambda_i + \mu_i)$$

Else

$$\inf_{w\in\mathbb{R}^d, b\in\mathbb{R}} L_{LHSVR}(w, b, \lambda, \mu) = -\infty$$

Linear Hard $\bar{\varepsilon}$-band SVR: Dual Form

$$\text{DLH} - \varepsilon\text{–band} - SVR \begin{cases} max \quad \dfrac{1}{2}\displaystyle\sum_{i,j=1}^{n}(\mu_i - \lambda_i)(\mu_j - \lambda_j)({x_i}^T x_i) - \varepsilon \sum_{i=1}^{n}(\mu_i + \lambda_i) + \sum_{i=1}^{n} y_i(\mu_i - \lambda_i) \\ s.à \qquad \displaystyle\sum_{i=1}^{n}(\mu_i - \lambda_i) = 0 \\ \qquad \lambda_i \geq 0, \mu_i \geq 0, \qquad i = 1, \ldots, n \end{cases}$$

Theorem:

1. If $\varepsilon > \varepsilon_{inf}$, then the DLH − ε–band − SVR problem has solutions
2. DLH − ε–band − SVR problem is convex quadratic programming
3. For any solution to the DLH − ε–band − SVR problem, $\lambda^* = (\lambda_1^*, \ldots, \lambda_n^*)$ and $\mu^* = (\mu_1^*, \ldots, \mu_n^*)$, if $\lambda^* \neq 0$ and $\mu^* \neq 0$, the solution to the LH − ε–band − SVR problem, (w^*, b^*), can be obtained in the following way
   - $w^* = \sum_{i=1}^{n}(\mu_i - \lambda_i)x_i$,
   - for any nonzero component $\lambda_j^*$of λ^*, $b^* = y_j - (w^*)^T x_j + \varepsilon$
   - Or for any nonzero component $\mu_j^*$of μ^*, $b^* = y_j - (w^*)^T x_j - \varepsilon$

Linear Hard **ε-band SVR** : Dual Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$
- Choose an approriate parameter $\varepsilon > 0$
- Contruct and solve the optimzation problem $\mathrm{DLH} - \varepsilon\text{–band} - SVR$ obtaining $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$ and $\mu^* = (\mu_1^*, \dots, \mu_n^*)$

$$
\mathrm{DLH} - \varepsilon\text{–band} - SVR \begin{cases} max \quad \dfrac{1}{2} \displaystyle\sum_{i,j=1}^{n} (\mu_i - \lambda_i)(\mu_j - \lambda_j)({x_i}^T x_i) - \varepsilon \sum_{i=1}^{n} (\mu_i + \lambda_i) + \sum_{i=1}^{n} y_i(\mu_i - \lambda_i) \\ s.\text{à} \qquad \displaystyle\sum_{i=1}^{n} (\mu_i - \lambda_i) = 0 \\ \qquad \lambda_i \geq 0, \mu_i \geq 0, \qquad i = 1, \dots, n \end{cases}
$$

if $\lambda^* \neq 0$ and $\mu^* \neq 0$, the solution to the problem, (w^*, b^*), can be obtained in the following way

- $w^* = \sum_{i=1}^{n} (\mu_i - \lambda_i) x_i$,
- for any nonzero component $\lambda_j^*$of λ^*, $b^* = y_j - (w^*)^T x_j + \varepsilon$
- Or for any nonzero component $\mu_j^*$of μ^*, $b^* = y_j - (w^*)^T x_j - \varepsilon$

$$h(x) = (w^*)^T x + b^*$$

- $\varepsilon < \varepsilon_{inf}$
- $D^+ : \varepsilon \leftarrow \varepsilon + \xi_i$
- $D^- : \varepsilon \leftarrow \varepsilon + \eta_i$

Linear Soft **ε-band Support Vector Regression(SVR)**

- Primal Problem

$$
\text{LS} - \varepsilon\text{–band} - SVR: \begin{cases} \min \quad \frac{1}{2} w^T w + C \sum_{i=1}^{n} (\xi_i + \eta_i) \\ y_i - (w^T x_i - b) \leq \varepsilon + \xi_i, \qquad i = 1, \dots, n \\ s.\text{à} \quad w^T x_i + b - y_i \quad \leq \varepsilon + \eta_i, \qquad i = 1, \dots, n \\ \xi_i \geq 0, \eta_i \geq 0, (w, b) \in \mathbb{R}^d \times \mathbb{R} \end{cases}
$$

Theorem:

- There exist solution to the $\text{LS} - \varepsilon\text{–band} - SVR$ problem w.r.t (w, b) and are not unique, In fact, when ε is large enough , , $(\bar{w}, \bar{b}, \bar{\xi}, \bar{\eta}) = (0, \bar{b}, 0, 0)$ are solutions, where $\bar{b}$ can take different values. Therefore $(\bar{w}, \bar{b}) = (0, \bar{b})$ are solutions w.r.t (w, b)
- The solution w.r.t. w is unique

Linear Soft **ε-band SVR** : Primal Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$
- **Choose an approriate parameter $\boldsymbol{\varepsilon > 0}$** and penalty parmeter
- Contruct and solve the optimzation problem $\mathrm{LS} - \varepsilon\text{–band} - SVR$ obtaining (w^*, b^*)

$$\mathrm{LS} - \varepsilon\text{–band} - SVR: \begin{cases} \quad min \quad \frac{1}{2} w^T w + C \sum_{i=1}^{n} (\xi_i + \eta_i) \\ \quad y_i - (w^T x_i - b) \leq \varepsilon + \xi_i, \qquad i = 1, \dots, n \\ s.à \quad w^T x_i + b - y_i \quad \leq \varepsilon + \eta_i, \qquad i = 1, \dots, n \\ \quad \xi_i \geq 0, \eta_i \geq 0, (w, b) \in \mathbb{R}^d \times \mathbb{R} \end{cases}$$

- Contruit the separating hyperplane $(w^*)^T x + b^*$ and the decision function is

$$h(x) = (w^*)^T x + b^*$$

Linear Soft ε-band SVR: Dual Form

In order to drive the dual problem of the primal we introduce lagrange function

$$L_{LSSVR}(w,b,\xi,\eta,\lambda,\mu,\alpha,\gamma) = f(w,\xi) + \sum_{i=1}^{n} \lambda_i f_i^+(w,b,\xi) + \sum_{i=1}^{n} \mu_i f_i^-(w,b,\eta) - \sum_{i=1}^{n} (\alpha_i \xi_i + \gamma_i \eta_i)$$

Where

- $f(w,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \eta_i)$
- $f_i^+(w,b,\xi) = y_i - (w^T x_i + b) - \varepsilon - \xi_i$
- $f_i^-(w,b,\eta) = w^T x_i + b - y_i - \varepsilon - \eta_i$

According to chapter 1, the dual problem should have a form of

$$DLH - \varepsilon\text{–band} - SVR \begin{cases} Max \quad g(\lambda,\mu,\alpha,\gamma) = \inf\limits_{w\in\mathbb{R}^d, b\in\mathbb{R}} L_{LSSVR}(w,b,\xi,\eta,\lambda,\mu,\alpha,\gamma) \\ s.t \qquad \lambda_i \geq 0, \mu_i \geq 0, \gamma_i \geq 0, \alpha_i \geq 0, \qquad i = 1,..,n \end{cases}$$

Linear Soft ε-band SVR: Dual Form

As $L_{LSSVR}(w, b, \xi, \eta, \lambda, \mu, \alpha, \gamma)$ is strictly convex quadratic Function of (w, b), its minimal value is achieved at (w, b, ξ, η)satisfying:

- $\nabla_w L_{LSSVR}(w, b, \xi, \eta, \lambda, \mu, \alpha, \gamma) = w - \sum_{i=1}^{n}(\lambda_i - \mu_i)x_i = 0 \Rightarrow w = \sum_{i=1}^{n}(\lambda_i - \mu_i)x_i$
- $\nabla_b L_{LSSVR}(w, b, \xi, \eta, \lambda, \mu, \alpha, \gamma) = \sum_{i=1}^{n}(\lambda_i - \mu_i) = 0$
- $\nabla_\xi L_{LSSVR}(w, b, \xi, \eta, \lambda, \mu, \alpha, \gamma) = CI_{n\times n} - \lambda - \alpha = 0$
- $\nabla_\eta L_{LSSVR}(w, b, \xi, \eta, \lambda, \mu, \alpha, \gamma) = CI_{n\times n} - \mu - \gamma = 0$

One has by substitution in $L_{LSSVR}(w, b, \xi, \lambda, \mu, \eta, \gamma)$:

If $\sum_{i=1}^{n}(\lambda_i - \mu_i) = 0$ then

$$\inf_{w\in\mathbb{R}^d, \xi\in\mathbb{R}^n b\in\mathbb{R}} L_{LSSVR}(w, b, \xi, \eta, \lambda, \mu, \alpha, \gamma) = \frac{1}{2}\sum_{i,j=1}^{n}(\mu_i - \lambda_i)(\mu_j - \lambda_j)({x_i}^T x_i) - \varepsilon\sum_{i=1}^{n}(\mu_i + \lambda_i) + \sum_{i=1}^{n} y_i(\mu_i - \lambda_i)$$

Else

$$\inf_{w\in\mathbb{R}^d, \xi\in\mathbb{R}^n b\in\mathbb{R}} L_{LSSVR}(w, b, \xi, \eta, \lambda, \mu, \alpha, \gamma) = -\infty$$

Linear Soft ε-band SVR: Dual Form

$$\mathrm{DLS} - \varepsilon\mathrm{-band} - SVR \begin{cases} Max & \frac{1}{2}\sum_{i,j=1}^{n}(\mu_i - \lambda_i)(\mu_j - \lambda_j)(xi^T xi) - \varepsilon\sum_{i=1}^{n}(\mu_i + \lambda_i) + \sum_{i=1}^{n} y_i(\mu_i - \lambda_i) \\ s.t & \sum_{i=1}^{n}(\mu_i - \lambda_i) = 0 \\ & C - \lambda_i - \mu_i = 0, i = 1, \dots, n \\ & \mu_i \geq 0, \lambda_i \geq 0 \; i = 1, \dots, n \end{cases}$$

- Dual can simplified to a problem only for a single variable λ by eliminating the variable μ and then rewritten as a minimization problem Dual $(Band - SVC)_\lambda$
- For $i = 1, \dots, n :\;\; C - \lambda_i - \mu_i = 0 \Leftrightarrow C - \lambda_i = \mu_i \geq 0 \Leftrightarrow C - \lambda_i \geq 0 \Leftrightarrow C \geq \lambda_i$

Linear Soft **ε-band SVR: Dual Form**

Dual form :

$$\text{DLS} - \varepsilon\text{–band} - SVR \begin{cases} max \quad \dfrac{1}{2}\displaystyle\sum_{i,j=1}^{n}(\mu_i - \lambda_i)(\mu_j - \lambda_j)({x_i}^T x_i) - \varepsilon \sum_{i=1}^{n}(\mu_i + \lambda_i) + \sum_{i=1}^{n} y_i(\mu_i - \lambda_i) \\ s.\text{à} \qquad \displaystyle\sum_{i=1}^{n}(\mu_i - \lambda_i) = 0 \\ \qquad C \geq \lambda_i \geq 0, C \geq \mu_i \geq 0, \qquad i = 1, \dots, n \end{cases}$$

Linear Soft **ε-band SVR** : Dual Algorithm

- Input: training set : $\{(x_i, y_i\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$
- **Choose an approriate parameter $\varepsilon > 0$** and penalty parameter $C > 0$
- Contruct and solve the optimzation problem $\mathrm{DLS} - \varepsilon\mathrm{-band} - SVR$ obtaining $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)$ and $\mu^* = (\mu_1^*, \dots, \mu_n^*)$

$$\mathrm{DLS} - \varepsilon\mathrm{-band} - SVR \begin{cases} min \quad \frac{1}{2}\sum_{i,j=1}^{n}(\mu_i - \lambda_i)(\mu_j - \lambda_j)(x_i{}^T x_i) - \varepsilon\sum_{i=1}^{n}(\mu_i + \lambda_i) + \sum_{i=1}^{n} y_i(\mu_i - \lambda_i) \\ s.\text{à} \qquad \sum_{i=1}^{n}(\mu_i - \lambda_i) = 0 \\ \qquad C \geq \lambda_i \geq 0, C \geq \mu_i \geq 0, \qquad i = 1, \dots, n \end{cases}$$

if $\lambda^* \neq 0$ and $\mu^* \neq 0$, the solution to the problem, (w^*, b^*), can be obtained in the following way

- $w^* = \sum_{i=1}^{n}(\mu_i - \lambda_i)x_i$,
- for any component $\lambda_j^* \in]0, C[$ of λ^*, $b^* = y_j - (w^*)^T x_j + \varepsilon$
- Or for any component $\mu_j^* \in]0, C[$ of μ^*, $b^* = y_j - (w^*)^T x_j - \varepsilon$

$$h(x) = (w^*)^T x + b^*$$