Support Vector machine with Random variables

Professor Abdellatif El Afia

Plan

1. Support Vector Classification with Random Variables
    1. Probabilistic Constraints Support Vector Classification
    2. Least Squares Probabilistic Support Vector Classification
2. Support Vector Regression with Random Variables
    1. Probabilistic Constraints Support Vector Regression

Support Vector machine

Support Vector Classification with Random Variables

Professor Abdellatif El Afia

Plan

1. Probabilistic Constraints Support Vector Classification
   1. Model Structure in Linear case
   2. Model Structure in NonLinear Case
2. Least Squares Probabilistic Support Vector Classification

Probabilistic Constraints $C - SVC$: Model Structure in Linear case

Frequently in pratical classification problems, training data cannot be observed precisely because of sampling errors, modeling errors or measurement errors. In this chapter, we investigate the SVC with unceratin input data.

- Suppoe that $\{(X_i, y_i)\}_{i=1}^{n}$ is the training set.
- $X_i = \left(X_i^1, \dots, X_i^d\right)^T$ is a random vector with $E(X_i) = \left(E(X_i^1), \dots, E(X_i^d)\right)^T$
- $y_i \in \{-1,1\}$
- $p_i \in [0,1]$, is the value of effect of the ith sample determination of the optimal hyperplane position.

$$C - SVC \begin{cases} Min & \dfrac{1}{2}\|w\|^2 + C\displaystyle\sum_{i=1}^{n}\xi_i \\ s.t & P(y_i(w^T X_i + b) \geq 1 - \xi_i) \geq p_i, i = 1, \dots, n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$$

- $\{(X_i, y_i)\}_{i=1}^{n}$
- $\{(x_i \pm \Delta x_i, y_i)\}_{i=1}^{n}$

Probabilistic Constraints $C - SVC$: Model Structure in Linear case

Since the optimization problem with probability inequality constraint is difficult to solve, we now derive sufficient conditions for this constraint and convert the optimisation problem into a solvable Quadratic Programming.

Theorem1

Let V be a random variable taking in the finite interval $[c, a]$. Then we have the follwing inequality

$$\left|P(V \leq v) - \frac{a - E(V)}{a - c}\right| \leq \frac{1}{2} + \frac{\left|v - \frac{c + a}{2}\right|}{a - c} \quad \forall v \in [c, a]$$

Corollary

$$\left|P(V \geq v) - \frac{E(V) - c}{a - c}\right| \leq \frac{1}{2} + \frac{\left|v - \frac{c + a}{2}\right|}{a - c}$$

Proof

- $P(V \leq v) = 1 - P(V \geq v) \Leftrightarrow P(V \leq v) - \frac{a-E(V)}{a-c} = 1 - P(V \geq v) - \frac{a-E(V)}{a-c}$
- $\Leftrightarrow P(V \leq v) - \frac{a-E(V)}{a-c} = -P(V \geq v) + \frac{E(V)-c}{a-c}$
- $\Leftrightarrow \left|P(V \leq v) - \frac{a-E(V)}{a-c}\right| = \left|-P(V \geq v) + \frac{E(V)-c}{a-c}\right| = \left|P(V \geq v) - \frac{E(V)-c}{a-c}\right| \leq \frac{1}{2} + \frac{\left|v-\frac{c+a}{2}\right|}{a-c}$

Probabilistic Constraints $C - SVC$: Model Structure in Linear case

Theorem2

For $a > 1$

$$y_i(w^T E(X_i) + b) \geq 2ap_i + 1 - \xi_i \Longrightarrow P(y_i(w^T X_i + b) \geq 1 - \xi_i) \geq p_i$$

Proof

For $a > 1$, $c = -a$, $\Longrightarrow c + a = 0$, $a - c = 2a$

If we put $V_i = y_i(w^T X_i + b)$, where $-a \leq V_i - \xi_i \leq V_i \leq V_i + \xi_i \leq a$

For $v_i \in [-a, a]$

- $\Longrightarrow \frac{E(V_i+\xi_i)+a}{2a} - \frac{1}{2} - \frac{\left|v_i - \frac{c+a}{2}\right|}{a-c} \leq P(V_i + \xi_i \geq v_i) \leq \frac{E(V_i+\xi_i)+a}{2a} + \frac{1}{2} + \frac{\left|v_i - \frac{c+a}{2}\right|}{a-c}$
- $\Longrightarrow \frac{E(V_i+\xi_i)+a}{2a} - \frac{1}{2} - \frac{|v_i|}{2a} \leq P(V_i + \xi_i \geq v_i) \leq \frac{E(V_i+\xi_i)+a}{2a} + \frac{1}{2} + \frac{|v_i|}{2a}$

Thus for $v_i = 1$, we have

$$\frac{E(V_i + \xi_i)}{2a} - \frac{1}{2a} \leq P(V_i + \xi_i \geq 1) \leq \frac{E(V_i + \xi_i)}{2a} + 1 + \frac{1}{2a}$$

Probabilistic Constraints $C - SVC$: Model Structure in Linear case

- $\Longrightarrow \frac{E(V_i+\xi_i)}{2a} - \frac{1}{2a} \leq P(y_i(w^T X_i + b) \geq 1 - \xi_i) \leq \frac{E(V_i+\xi_i)}{2a} + 1 + \frac{1}{2a}$
- $\Longrightarrow \frac{E(V_i)+\xi_i}{2a} - \frac{1}{2a} \leq P(y_i(w^T X_i + b) \geq 1 - \xi_i) \leq \frac{E(V_i)+\xi_i}{2a} + 1 + \frac{1}{2a}$

Since $E(V_i) = y_i(w^T E(X_i) + b) \geq 2ap_i + 1 - \xi_i$

- $\Longrightarrow E(V_i) - 1 + \xi_i \geq 2ap_i$
- $\Longrightarrow \frac{E(V_i)+\xi_i}{2a} - \frac{1}{2a} \geq p_i$
- $\Longrightarrow P(y_i(w^T X_i + b) \geq 1 - \xi_i) \geq p_i$

So that

$$\frac{E(V_i) + \xi_i}{2a} + 1 + \frac{1}{2a} = \frac{E(V_i) + \xi_i}{2a} - \frac{1}{2a} + 1 + \frac{1}{2a} \geq p_i + 1 + \frac{1}{2a} \geq 1$$

- $\Longrightarrow P(y_i(w^T X_i + b) \geq 1 - \xi_i) \leq 1$

- $P(y_i(w^T X_i + b) \geq 1 - \xi_i) \geq p_i$

Theorem2

For $a > 1$

$$y_i(w^T E(X_i) + b) \geq 2ap_i + 1 - \xi_i \Longrightarrow P(y_i(w^T X_i + b) \geq 1 - \xi_i) \geq p_i$$

- $\begin{cases} Min & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ s.t & y_i(w^T E(X_i) + b) \geq 2ap_i + 1 - \xi_i, i = 1,..,n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$

- $\Longrightarrow \begin{cases} Min & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ s.t & P(y_i(w^T X_i + b) \geq 1 - \xi_i) \geq p_i, i = 1,..,n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$

Probabilistic Constraints $C-SVC$: Model Structure in Linear case

The optimal separating hyperplane can be obtained by solving the follwing optimization problem

$$C-SVC\begin{cases} Min & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ s.t & y_i(w^T E(X_i)+b) \geq 2ap_i + 1 - \xi_i, i = 1,..,n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$$

Similar to the standard $C-SVC$, the optimization problem of Probabilistic Constraints $C-SVC$ can transformed into its dual problem

$$D-SVC:\begin{cases} Max & g(\lambda) = \sum_{i=1}^{n}\lambda_i(2ap_i+1) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_j\lambda_i y_j y_i\big(E(X_j)^T E(X_i)\big) \\ s.t & \sum_{i=1}^{n}\lambda_i y_i = 0 \\ & C \geq \lambda_i \geq 0 \ i = 1,\dots,n \end{cases}$$

Probabilistic Constraints $C - SVC$: Model Structure in Linear case

If the optimum Lagrange multipliers denotes by $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)^T$, we may compute the optimum weight vector $w^*$and bias b^*respectively by using the following equations:

- $w^* = \sum_{i=1}^{n} \lambda_i^* y_i E(X_i)$
- $b^* = \begin{cases} 2ap_i + 1 - (w^*)^T E(X_i) & if \;\; y_i = 1, \;\; \lambda_i^* \in]0, C[\\ -2ap_i - 1 - (w^*)^T E(X_i) & if \;\; y_i = -1, \;\; \lambda_i^* \in]0, C[\end{cases}$

If probability function of $X_i^1, \dots, X_i^d$are unknown then $E(X_i^1), \dots, E(X_i^d)$ are unknown. In this case using statistical methods, we apply the plug-in estimators and finally we should change optimization problem $D - SVC$.

For each input random vector $\mathrm{E}(X_i) = \left(E(X_i^1), \dots, E(X_i^d)\right)^T$, we randomly generatee n_i samples x_{ik},

$$x_{ik} = \left(x_{ik}^1, \dots, x_{ik}^d\right)^T k = 1, \dots, n_i$$

according to the $X_i^1, \dots, X_i^d$. Afterwards we apply the plug-in estimator $\bar{x}_i$

$$\bar{x}_i = \left(\frac{1}{n_i}\sum_{k=1}^{n_i} x_{ik}^1, \dots, \frac{1}{n_i}\sum_{k=1}^{n_i} x_{ik}^d\right)^T = \left(\bar{x}_i^1, \dots, \bar{x}_i^d\right)^T$$

Probabilistic Constraints $C - SVC$: Model Structure in Linear case

Instead of $E(X_i)$ into the optimization problem $D - SVC$. Therefore, the optimal seperating hyperplane can obtained by solving the following optimization problem:

$$D - SVC: \begin{cases} Max \quad g(\lambda) = \sum_{i=1}^{n} \lambda_i (2ap_i + 1) - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_j \lambda_i y_j y_i \left({\bar{x}_i}^T \bar{x}_j \right) \\ s.t \quad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ C \geq \lambda_i \geq 0 \; i = 1, \dots, n \end{cases}$$

If the optimum Lagrange multipliers denotes by $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)^T$, Then the estimation of the optimum weight vector $w^*$and bias b^*respectively by using the following equations:

- $\hat{w}^* = \sum_{i=1}^{n} \lambda_i^* y_i \bar{x}_i$
- $\hat{b}^* = \begin{cases} 2ap_i + 1 - (\hat{w}^*)^T \bar{x}_i \quad if \;\; y_i = 1, \quad \lambda_i^* \in]0, C[\\ -2ap_i - 1 - (\hat{w}^*)^T \bar{x}_i \quad if \;\; y_i = -1, \;\; \lambda_i^* \in]0, C[\end{cases}$

Probabilistic Constraints $C-SVC$: Model Structure in NonLinear case

A nonlinear transformation ϕ is used to transform data points from the input space of dimension d into a feature space having dimension m. The nonlinear mapping is denoted by $\phi: \mathbb{R}^d \rightarrow \mathbb{R}^m$. Suppoe that $\{(\phi(X_i), y_i \in \{-1,1\})\}_{i=1}^{n}$ is the training set such that:

- $\phi(X_i) = (\phi^1(X_i), \dots, \phi^m(X_i))^T$ is a random vector
- $E(\phi(X_i)) = (E(\phi^1(X_i)), \dots, E(\phi^m(X_i)))^T$ is its expectation
- $p_i \in [0,1]$ is the value of effect of the ith sample determination of the optimal hyperplane position.

$$C-SVCNS \begin{cases} Min & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ s.t & P(y_i(w^T\phi(X_i)+b) \geq 1-\xi_i) \geq p_i, i=1,..,n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$$

Theorem2:

For $a > 1$ $y_i(w^T E(\phi(X_i)) + b) \geq 2ap_i + 1 - \xi_i \Longrightarrow P(y_i(w^T\phi(X_i)+b) \geq 1-\xi_i) \geq p_i$

Probabilistic Constraints $C - SVC$: Model Structure in NonLinear case

Then we have $C - SVCNS$ problem:

- $$\begin{cases} Min & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ s.t & y_i(w^T E(\phi(X_i)) + b) \geq 2ap_i + 1 - \xi_i, i = 1,..,n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$$

and its dual problem

- $$\begin{cases} Max & g(\lambda) = \sum_{i=1}^{n}\lambda_i(2ap_i + 1) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_j\lambda_i y_j y_i \left(E(\phi(X_j))^T E(\phi(X_i))\right) \\ s.t & \sum_{i=1}^{n}\lambda_i y_i = 0 \\ & C \geq \lambda_i \geq 0 \; i = 1, \dots, n \end{cases}$$

If $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)^T$, the optimum weight vector $w^*$and bias b^*respectively by :

- $w^* = \sum_{i=1}^{n}\lambda_i^* y_i E(\phi(X_i))$
- $b^* = \begin{cases} 2ap_i + 1 - (w^*)^T E\big(\phi(X_i)\big) & if \;\; y_i = 1, \;\; \lambda_i^* \in]0, C[\\ -2ap_i - 1 - (w^*)^T E\big(\phi(X_i)\big) & if \;\; y_i = -1, \; \lambda_i^* \in]0, C[\end{cases}$

Probabilistic Constraints $C - SVC$: Model Structure in NonLinear case

Corollary:

For each $X_i = (X_i^1, \dots, X_i^d)^T$, we randomly generate n_i samples

$$x_{ik} = \left(x_{ik}^1, \dots, x_{ik}^d\right)^T k = 1, \dots, n_i$$

Afterwards we apply the plug-in estimator $\hat{E}\left(\phi(X_i)\right)$,

- $\hat{E}\left(\phi(X_i)\right) = \left(\hat{E}(\phi^1(X_i)), \dots, \hat{E}(\phi^m(X_i))\right)^T = \left(\frac{1}{n_i}\sum_{k=1}^{n_i} \phi^1(x_{ik}), \dots, \frac{1}{n_i}\sum_{k=1}^{n_i} \phi^m(x_{ik})\right)^T$
- $\Longrightarrow \hat{E}\left(\phi(X_i)\right) = \frac{1}{n_i}\sum_{k=1}^{n_i}(\phi^1(x_{ik}), \dots, \phi^m(x_{ik}))^T = \frac{1}{n_i}\sum_{k=1}^{n_i} \phi(x_{ik})$

Instead of $\hat{E}\left(\phi(X_i)\right)$ into the $C - SVCNS$, then we have :

- $$\begin{cases} Min & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \\ s.t & y_i(\frac{1}{n_i}\sum_{k=1}^{n_i} w^T\phi(x_{ik}) + b) \geq 2ap_i + 1 - \xi_i, i = 1, .., n \\ & \xi_i \geq 0, w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}$$

- $X_i = (X_i^1, \dots, X_i^d)^T$
- $x_i = (x_i^1 \pm \Delta x_i^1, \dots, x_i^d \pm \Delta x_i^d)^T$
- $x_i^k \pm \Delta x_i^k \in [a_k, b_k]$
- $\{(x_i \pm \Delta x_i, y_i)\}_{i=1}^n$

Probabilistic Constraints $C-SVC$: Model Structure in NonLinear case

and its dual $DC-SVCNS$

- $$\begin{cases} Max \quad g(\lambda) = \sum_{i=1}^{n} \lambda_i (2ap_i + 1) - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} \frac{\lambda_j \lambda_i y_j y_i}{n_i n_j} (\phi(x_{ik}))^T \phi(x_{ik}) \\ s.t \quad \sum_{i=1}^{n} \lambda_i y_i = 0 \\ \quad C \geq \lambda_i \geq 0 \; i = 1, \dots, n \end{cases}$$

If the optimum Lagrange multipliers denotes by $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)^T$, Then the estimation of the optimum weight vector $w^*$and bias b^*respectively by using the following equations:

- $\widehat{w}^* = \sum_{i=1}^{n} \lambda_i^* y_i \widehat{E}\big(\phi(X_i)\big)$
- $\widehat{b}^* = \begin{cases} 2ap_i + 1 - (\widehat{w}^*)^T \widehat{E}\big(\phi(X_i)\big) & if \;\; y_i = 1, \;\; \lambda_i^* \in]0, C[\\ -2ap_i - 1 - (\widehat{w}^*)^T \widehat{E}\big(\phi(X_i)\big) & if \;\; y_i = -1, \;\; \lambda_i^* \in]0, C[\end{cases}$

Probabilistic Constraints $C - SVC$: TP1

Let $(X_i, y_i), i = 1, \dots, n$ such that: $n \in \{20,50,100\}$

- $X_i = (X_i^1, X_i^2)^T, i = 1, \dots, n$:
- $y_i = \begin{cases} 1 & i = 1, \dots, m \\ -1 & i = m+1, \dots, n \end{cases}$

- For $i = 1, \dots, m,$ you randomply generate $n_i = 30$ samples for any of
  - $X_i^1 \rightsquigarrow \mathcal{N}(\mu = 5, \sigma = 2)$
  - $X_i^2 \rightsquigarrow \mathcal{N}(\mu = 3, \sigma = 1)$
- For $i = m+1, \dots, n,$ you randomply generate $n_i = 30$ samples for any of
  - $X_i^1 \rightsquigarrow \mathcal{N}(\mu = 5, \sigma)$
  - $X_i^2 \rightsquigarrow \mathcal{N}(\mu = 2, \sigma)$
- Compute $\bar{x}_i = \left(\frac{1}{n_i}\sum_{k=1}^{n_i} x_{ik}^1, \frac{1}{n_i}\sum_{k=1}^{n_i} x_{ik}^2\right)^T = \left(\bar{x}_i^1, \bar{x}_i^2\right)^T,$
- Consider $a = 2, p_i = 0{,}9$ and you choose appropriate $C \in \{5,20,50,100,150,200,250,300\}$
- Solve the optimization problem ($D - SVC$ or C$-SVC$) and contruit :
  - the separating hyperplane $h_{w^*,b^*}(x) = (w^*)^T x + b^*$
  - the decision function is $h_S(x) = sign\left(h_{w^*,b^*}(x)\right) \rightarrow L_S(h_S) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{h_{w^*,b^*}(x_i) \neq y_i\}}$

Probabilistic Constraints $C - SVC$: TP2

Let $(X_i, y_i), i = 1, \dots, n$ such that: $n \in \{20{,}50{,}100\}$

- $X_i = (X_i^1, X_i^2)^T, i = 1, \dots, n$:
- $y_i = \begin{cases} 1 & i = 1, \dots, m \\ -1 & i = m+1, \dots, n \end{cases}$

- For $i = 1, \dots, m,$ you randomply generate $n_i = 30$ samples for any of
  - $X_i^1 \rightsquigarrow \mathcal{U}(a_i, b_i)$ where $a_i \in]1,2[, b_i \in]2,3[$
  - $X_i^2 \rightsquigarrow \mathcal{U}(c_i, d_i)$ where $c_i \in]2,3[, d_i \in]3,4[$
- For $i = m+1, \dots, n,$ you randomply generate $n_i = 30$ samples for any of
  - $X_i^1 \rightsquigarrow \mathcal{U}(a_i', b_i')$ where $a_i' \in]2,3[, b_i' \in]3,4[$
  - $X_i^2 \rightsquigarrow \mathcal{U}(c_i', d_i')$ where $c_i' \in]1,2[, d_i' \in]2,3[$
- Compute $\bar{x}_i = \left(\frac{1}{n_i}\sum_{k=1}^{n_i} x_{ik}^1, \frac{1}{n_i}\sum_{k=1}^{n_i} x_{ik}^2\right)^T = \left(\bar{x}_i^1, \bar{x}_i^2\right)^T$,
- Consider $C = 200\ a = 2,$ and $p_i \in \{0{,}9; 0{,}8{,}0{,}7; 0{,}6; 0{,}5; 0{,}4; 0{,}3;\ 0{,}2; 0{,}1\}$
- Solve the optimization problem ($D - SVC$ or C$-SVC$) and contruit :
  - the separating hyperplane $h_{w^*,b^*}(x) = (w^*)^T x + b^*$
  - the decision function is $h_S(x) = sign\left(h_{w^*,b^*}(x)\right) \to L_S(h_S) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{h_{w^*,b^*}(x_i) \neq y_i\}}$

Least Squares Probabilistic Support Vector Classification

We introduce a least squares version to the Probabilistic Support Vector Classification

$$
LS-PSVC \begin{cases} Min & \frac{1}{2}\left(\|w\|^2 + C\sum_{i=1}^{n}(\xi_i)^2\right) \\ s.t & y_i\left(\frac{1}{n_i}\sum_{k=1}^{n_i} w^T\phi(x_{ik}) + b\right) = 2ap_i + 1 - \xi_i, i = 1,..,n \\ & w \in \mathbb{R}^d, b \in \mathbb{R} \end{cases}
$$

And its dual ?

Support Vector machine

Support Vector Regression with Random Variables

Professor Abdellatif El Afia

Plan

1. Probabilistic Constraints Support Vector Regression
    1. Model Structure in Linear case
    2. Model Structure in NonLinear Case

Probabilistic Constraints Support Vector Regression

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

And its dual

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

Probabilistic Constraints Support Vector Regression

Frenquently in pratical regression models, training data, $\{(x_i, y_i)\}_{i=1}^{n}$, containing input and output data cannot be observed precisely because of sampling errors, thus usually they are presented by random variables. In order to achieve robustness, the constraints in SVR problem must be replaced with probability constraints.

Probablistic constraints SVR finds the optimal hyperplane regression $h_{w,b}(x) = w^T x + b$.

In this section we deal with randomized output Y_i and randomized bias B such that :

- $Y_i \sim \mathcal{U}(l_i, u_i) \Longrightarrow f_{Y_i}(y_i) = \begin{cases} \frac{1}{u_i - l_i} & if \quad y_i \in (l_i, u_i) \\ 0 & otherwise \end{cases}$

- $B \sim \mathcal{U}(l'_i, u'_i) \Longrightarrow f_B(b) = \begin{cases} \frac{1}{u'_i - l'_i} & if\ b \in (l'_i, u'_i) \\ 0 & otherwise \end{cases}$

Also we suppose that Y_i and B are independent together, then $f_{Y_i,B}(y_i, b) = f_{Y_i}(y_i) f_B(b)$

Model Structure in Linear case

In the proposed algorithm, optimal hyperplane regression can obtained by solving the following optimization problem

$$\text{LS}-\varepsilon\text{–band}-SVR: \begin{cases} \quad min \quad \dfrac{1}{2}w^T w + C\displaystyle\sum_{i=1}^{n}(\xi_i + \eta_i) \\ \quad P(\mathrm{Y}_i - w^T x_i - B \le \varepsilon + \xi_i) \ge p, \qquad i = 1, \dots, n \\ s.\text{à} \quad P(w^T x_i + \mathrm{B} - \mathrm{Y}_i \le \varepsilon + \eta_i) \ge p, \qquad i = 1, \dots, n \\ \quad \xi_i \ge 0, \eta_i \ge 0, w \in \mathbb{R}^d \end{cases}$$

Where $p \in [0,1]$

The optimization problem with inequality constraints is difficult to solve we now convert the optimization problem a solvable quadratic problem using the probability theory

Model Structure in Linear case: Probability Theory

- $P(\mathrm{Y}_i - (w^T x_i + B) \leq \varepsilon + \xi_i) = P(\mathrm{Y}_i - B \leq w^T x_i + \varepsilon + \xi_i)$
- $P(\mathrm{Y}_i - B \leq w^T x_i + \varepsilon + \xi_i) = \int_{l_i'}^{u_i'} \int_{l_i}^{w^T x_i + \varepsilon + \xi_i + b} f_{Y_i}(y_i) f_B(b) dy_i db$
- $\int_{l_i'}^{u_i'} (\int_{l_i}^{w^T x_i + \varepsilon + \xi_i + b} \frac{1}{u_i - l_i} dy_i) \frac{1}{u_i' - l_i'} db = \int_{l_i'}^{u_i'} \frac{w^T x_i + \varepsilon + \xi_i + b - l_i}{(u_i - l_i)(u_i' - l_i')} db = \frac{\boldsymbol{w^T x_i + \varepsilon + \xi_i + b - l_i + \frac{1}{2}(u_i' + l_i')}}{\boldsymbol{(u_i - l_i)}}$
- $P(\mathrm{Y}_i - w^T x_i - B \leq \varepsilon + \xi_i) = \frac{\boldsymbol{w^T x_i + \varepsilon + \xi_i - l_i + \frac{1}{2}(u_i' + l_i')}}{\boldsymbol{(u_i - l_i)}}$

And

- $P(w^T x_i + \mathrm{B} - \mathrm{Y}_i \leq \varepsilon + \eta_i) = P(\mathrm{B} - \mathrm{Y}_i \leq -w^T x_i + \varepsilon + \eta_i)$
- $P(\mathrm{B} - \mathrm{Y}_i \leq -w^T x_i + \varepsilon + \eta_i) = \int_{l_i'}^{u_i'} \int_{b + w^T x_i - \varepsilon - \eta_i}^{l_i} f_{Y_i}(y_i) f_B(b) dy_i db$
- $\int_{l_i'}^{u_i'} (\int_{b + w^T x_i - \varepsilon - \eta_i}^{l_i} \frac{1}{u_i - l_i} dy_i) \frac{1}{u_i' - l_i'} db = \int_{l_i'}^{u_i'} \frac{l_i - w^T x_i + \varepsilon + \eta_i - b}{(u_i - l_i)(u_i' - l_i')} db = \frac{l_i - w^T x_i + \varepsilon + \eta_i - \frac{1}{2}(u_i' + l_i')}{(u_i - l_i)}$
- $\Longrightarrow P(\mathrm{B} - \mathrm{Y}_i \leq -w^T x_i + \varepsilon + \eta_i) = \frac{\boldsymbol{l_i - w^T x_i + \varepsilon + \eta_i - \frac{1}{2}(u_i' + l_i')}}{\boldsymbol{(u_i - l_i)}}$

Model Structure in Linear case

Then LS − ε−band − SVR problem can be transformed into the following form

- $$\begin{cases} \min \quad \frac{1}{2} w^T w + C \sum_{i=1}^{n} (\xi_i + \eta_i) \\ w^T x_i + \frac{1}{2}(u_i' + l_i') - l_i + \varepsilon + \xi_i \geq p(u_i - l_i), \quad i = 1, \dots, n \\ s.à \quad l_i - w^T x_i - \frac{1}{2}(u_i' + l_i') + \varepsilon + \eta_i \geq p(u_i - l_i), \quad i = 1, \dots, n \\ \xi_i \geq 0, \eta_i \geq 0, w \in \mathbb{R}^d \end{cases}$$

And its dual

- $$\begin{cases} \max \quad \frac{1}{2} \sum_{i,j=1}^{n} (\mu_i - \lambda_i)(\mu_j - \lambda_j)({x_i}^T x_j) - \varepsilon \sum_{i=1}^{n} (\mu_i + \lambda_i) + \sum_{i=1}^{n} \lambda_i (p u_i + (1-p) l_i) - \sum_{i=1}^{n} \mu_i ((1-p) u_i + p l_i) \\ s.à \quad \sum_{i=1}^{n} (\mu_i - \lambda_i) = 0 \\ C \geq \lambda_i \geq 0, C \geq \mu_i \geq 0, \ i = 1, \dots, n \end{cases}$$

Model Structure in Linear case

We know that $E(B) = \frac{1}{2}(u_i' + l_i') = \mu_B$,

we represent optimal value of μ_B by $\hat{\mu}_B$ *and* optimal value of w by $\widehat{w}$

If the optimum Lagrange multipliers denotes by $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)^T$ and $\mu^* = (\mu_1^*, \dots, \mu_n^*)^T$ we may compute the optimum weight vector $\widehat{w}$ and bias $\hat{\mu}_B$ respectively by using the following equations:

- $\widehat{w} = \sum_{i=1}^{n}(\lambda_i^* - \mu_i^*)x_i$
- $\begin{cases} \hat{\mu}_B = p(u_i - l_i) - \widehat{w}^T x_i - l_i + \varepsilon & For\ \lambda_i^* \in (0, C), i = 1, \dots, n \\ \hat{\mu}_B = p(u_i - l_i) - l_i - \widehat{w}^T x_i + \varepsilon & For\ \mu_i^* \in (0, C), i = 1, \dots, n \end{cases}$

Thus, we can find optimal hyperplane regression as

$$\hat{h}_{\widehat{w}, \hat{\mu}_B}(x) = \sum_{i=1}^{n}(\lambda_i - \mu_i)x_i^T x + \hat{\mu}_B$$

Where $\hat{h}_{\widehat{w}, \hat{\mu}_B}$ is estimation of $E\big(h_{w,B}(x)\big) = E(w^T x + B)$

Model Structure in Linear case TP1

- $C = 100, \varepsilon = 0.1$ and $\text{p} = 0.99$
- Generate randomly $x_i = (x_i^1, x_i^2)$ for $i = 1, \ldots, 20$ from uniform distribution on $(0,10)$
- Compute the corresponding, l_i et u_i, with $\mu_{0B} = 5$, δ_i is a random point on $(0,1)$, and $w_0 \in \{(0.6,1.4), (1.4, 1)\}$
  - $l_i = (w_0)^T x_i + \mu_{0B} - \delta_i$
  - $u_i = (w_0)^T x_i + \mu_{0B} + \delta_i$
- Add to x_i, l_i and u_i a noise=$\mathcal{N}(\mu = 0, \Sigma \in (0,1))$
- Generate $y_i \in \mathcal{U}(l_i, u_i)$

Model Structure in NonLinear case

In the proposed algorithm, optimal hyperplane regression can obtained by solving the following optimization problem

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

Where $p \in [0,1]$

The optimization problem with inequality constraints is difficult to solve we now convert the optimization problem a solvable quadratic problem using the probability theory

Model Structure in NonLinear case

Then LS – ε–band – SVR problem can be transformed into the following form

- $$\begin{cases} \quad\quad\quad\quad\quad\quad min \quad \frac{1}{2}w^T w + C\sum_{i=1}^{n}(\xi_i + \eta_i) \\ \quad\quad w^T\phi(x_i) + \frac{1}{2}(u_i' + l_i') - l_i + \varepsilon + \xi_i \geq p(u_i - l_i), \quad i = 1, \dots, n \\ s.à \quad l_i - w^T\phi(x_i) - \frac{1}{2}(u_i' + l_i') + \varepsilon + \eta_i \geq p(u_i - l_i), \quad i = 1, \dots, n \\ \quad\quad\quad\quad\quad\quad \xi_i \geq 0, \eta_i \geq 0, w \in \mathbb{R}^d \end{cases}$$

And its dual

- $$\begin{cases} max \quad \frac{1}{2}\sum_{i,j=1}^{n}(\mu_i - \lambda_i)(\mu_j - \lambda_j)(\phi(x_i)^T\phi(x_j)) - \varepsilon\sum_{i=1}^{n}(\mu_i + \lambda_i) + \sum_{i=1}^{n}\lambda_i(pu_i + (1-p)l_i) - \sum_{i=1}^{n}\mu_i((1-p)u_i + pl_i) \\ \quad\quad\quad\quad\quad\quad\quad\quad \sum_{i=1}^{n}(\mu_i - \lambda_i) = 0 \\ \quad\quad s.à \\ \quad\quad\quad\quad\quad\quad\quad\quad C \geq \lambda_i \geq 0, C \geq \mu_i \geq 0, \; i = 1, \dots, n \end{cases}$$

Model Structure in NonLinear case

If the optimum Lagrange multipliers denotes by $\lambda^* = (\lambda_1^*, \ldots, \lambda_n^*)^T$ and $\mu^* = (\mu_1^*, \ldots, \mu_n^*)^T$ we may compute the optimum weight vector $\widehat{w}$ and bias $\hat{\mu}_B$ respectively by using the following equations:

- $\widehat{w} = \sum_{i=1}^{n} (\lambda_i^* - \mu_i^*)\phi(x_i)$
- $\begin{cases} \hat{\mu}_B = p(u_i - l_i) - \widehat{w}^T\phi(x_i) - l_i + \varepsilon & For\ \lambda_i^* \in (0, C), i = 1, \ldots, n \\ \hat{\mu}_B = p(u_i - l_i) - l_i - \widehat{w}^T\phi(x_i) + \varepsilon & For\ \mu_i^* \in (0, C), i = 1, \ldots, n \end{cases}$

Thus, we can find optimal hyperplane regression as

$$\widehat{h}_{\widehat{w},\widehat{\mu}_B}(x) = \sum_{i=1}^{n} (\lambda_i - \mu_i)\phi(x_i)\phi(x) + \hat{\mu}_B$$

Where $\widehat{h}_{\widehat{w},\widehat{\mu}_B}$ is estimation of $E\left(h_{w,B}(x)\right) = E(w^T\phi(x) + B)$