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

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

And its dual

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

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

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

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 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

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

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 $E\left(h_{w,B}(x)\right) = E(w^T x + B)$

Model Structure in Linear case TP1

- $C = 100, \varepsilon = 0.1$ and $\mathrm{p} = 0.99$
- Generate randomly $x_i = (x_i^1, x_i^2)$ for $i = 1, \dots, 20$ 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

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

And its dual

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

Model Structure in NonLinear case

If the optimum Lagrange multipliers denotes by $\lambda^* = (\lambda_1^*, \dots, \lambda_n^*)^T$ 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:

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

Thus, we can find optimal hyperplane regression as

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

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