ITESO, Universidad Jesuita de Guadalajara

Maestría de Ciencia de Datos

Optimización Convexa

Tarea 8: SVR

Estudiante: Daniel Nuño
Profesor: Dr. Juan Diego Sanchez Torres
Fecha entrega: 16 de marzo, 2022

Introduction

For the case of SVM for regression, or SVR, let the set $D = (x_1, y_1), \dots, (x_N, y_N)$, where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}$. Let $\varphi : X \to \mathcal{F}$ be the function that makes each input point x correspond a point in the feature space $\mathcal{F}$, where $\mathcal{F}$ is a Hilbert space. This feature space can be of high dimension or even infinite. However, is common to define $X = \mathbb{R}^n$ and $\mathcal{F} = \mathbb{R}^m$. In this form, the approximating function, namely the model, has the form $\hat{y}_k = f(x_k) = w^T \varphi(x_k) + b$ with $w \in \mathbb{R}^m$ and $b \in \mathbb{R}$.

Commonly, the first approach for solving the SVR is the L_1^ϵ formulation. The following problem statement considers such a regression problem as a convex optimization problem.

Activities

First, the following case considers the application of L_1^ϵ-SVR.

Problem 1: The L_1^ϵ SVR

Consider the following optimization problem:

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \mathcal{P}_\epsilon(w,b,\xi,\xi^*) = & \frac{1}{2} w^T w + c \sum_{k=1}^{N} (\xi_k + \xi_k^*) \\ \text{s. t. } & y_k - w^T \varphi(x_k) - b \le \epsilon + \xi_k,\ k = 1, \dots, N \\ & w^T \varphi(x_k) + b - y_k \le \epsilon + \xi_k^*,\ k = 1, \dots, N \\ & \xi_k, \xi_k^* \ge 0,\ k = 1, \dots, N \end{aligned} \tag{1}$$

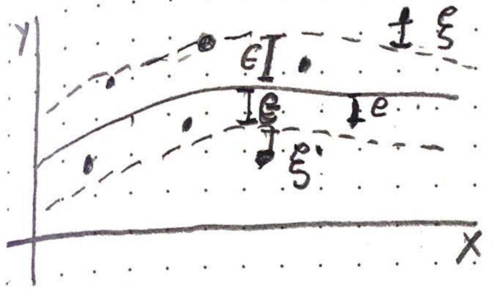
where $\varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ and the regularization parameter $c > 0$ determines the balance between the regularity of f and the quantity up to which we tolerate deviations more significant than ϵ. Consider ξ_k and ξ_k^* as slack variables that control the error between the prediction $\hat{y}_k$ and the k-th sample y_k.

For this case, present a complete development of the problem (1), by using the KKT optimality approach.

9 marzo, 22

Problema 1: L_1^{ϵ} SVR

considera la siguiente gráfica en 2 dimensiones en la que se quiere plantear una regresión lineal



$y_k - W^T\varphi(x_k) - b \Rightarrow$ Arriba

$W^T\varphi(x_k) + b - y_k \Rightarrow$ Abaja

ϵ es la diferencia entre la linea solida y la linea punteada.

error es el valor entre cada punto y la linea solida.

ξ es el error en cada punto por afuera de la linea punteada

Los vectores de soporte son los que estan por afuera del tubo.

El problema se vuelve relevante solo a los valores que estan por afuera del tubo.

primero definimos Lagrange como

$L(W,b,\xi,\xi^*,\alpha,\alpha^*,\eta,\eta^*)$ siendo $\alpha,\alpha^*,\eta,\eta^*$ los valores de lagrange

$$= \frac{1}{2}W^TW + C\sum(\xi_k + \xi_k^*) - \sum\alpha_k(\epsilon + \xi_k - y_k - W^T\varphi(x_k) + b)$$

$$- \sum\alpha_k^*(\epsilon + \xi_k^* - y_k - W^T\varphi(x_k) - b + y_k)$$

$$- \sum\eta_k\xi_k$$

$$- \sum\eta_k^*\xi_k^*$$

Usamos KKT1, derivamos respecto las variables del primal e igualamos a cero

$$\nabla W = W - \sum\alpha_k\varphi(x_k) + \sum\alpha_k^*\varphi(x_k) = 0$$

$$W = \sum\alpha_k\varphi(x_k) - \sum\alpha_k^*\varphi(x_k)$$

9 marzo, 22

$$\frac{\partial L}{\partial b} = \sum \alpha_k + \sum \alpha_k^* = 0$$

$$\sum (\alpha_k + \alpha_k^*) = 0$$

$$\frac{\partial L}{\partial \xi_k^*} = C - \alpha_k^* - \eta_k^* = 0$$

$$C - \alpha_k = \eta_k$$

$$\frac{\partial L}{\partial \xi_k} = C - \alpha_k - \eta_k = 0$$

$$C - \alpha_k^* = \eta_k^*$$

Ahora simplificamos y escribimos el dual

$$L = \frac{1}{2} W^T W + C \sum (\xi_k + \xi_k^*) - \epsilon \sum (\alpha_k + \alpha_k^*) - \sum \alpha_k \xi_k$$
$$- \sum \alpha_k^* \xi_k^* + \sum (\alpha_k - \alpha_k^*) y_k - \sum (\alpha_k - \alpha_k^*) W^T \varphi(x_k)$$
$$- b \sum (\alpha_k + \alpha_k^*) - C \sum (\xi_k + \xi_k^*) + \sum \alpha_k \xi_k + \sum \alpha_k^* \xi_k^*$$

Lagrange Dual se escribe como

$$D = \frac{1}{2} \sum_{l=1}^{N} \sum_{k=1}^{N} (\alpha_l - \alpha_l^*)(\alpha_k - \alpha_k^*) \varphi(x_l)^T \varphi(x_k) + \sum (\alpha_k - \alpha_k^*) y_k$$
$$- \epsilon \sum (\alpha_k + \alpha_k^*)$$

siendo el dual que busca maximizar α_k y α_k^*

$$\max_{\alpha_k, \alpha_k^*} D = \frac{1}{2} \sum \sum (\alpha_l - \alpha_l^*)(\alpha_k - \alpha_k^*) \varphi(x_l)^T \varphi(x_k)^T + \sum (\alpha_k + \alpha_k^*) y_k - \epsilon \sum (\alpha_k + \alpha_k^*)$$

$$\text{s.t. } \sum (\alpha_k + \alpha_k^*) = 0 \qquad 0 \leq \alpha_k < C$$

$$\eta_k = C - \alpha_k, \quad \eta_k \geq 0$$

$$C - \alpha_k \geq 0$$

$$C \geq 0$$

9 marzo, 22.

KKT2 son las restricciones del primal

$$y_k - W^T \phi(x_k)(x_k) - b \leq \epsilon + \xi_k$$

$$W^T \phi(x_k) + b - y_k \leq \epsilon + \xi_k^*$$

$$\xi_k \geq 0$$

$$\xi_k^* \geq 0$$

KKT3 son las restricciones del dual

$$\sum_{k=1}^{N} (\alpha_k - \alpha_k^*) = 0$$

$$0 \leq \alpha_k \leq C$$

$$0 \leq \alpha_k^* \leq C$$

KKT4 son las restricciones KKT2 y KKT3 más la condición complementaria

$\eta_k \xi_k = 0$ condiciones complementarias

$$\eta_k^* \xi_k^* = 0$$

$$\alpha_k (y_k - W^T \phi(x_k) - b - \epsilon - \xi_k) = 0$$

$$\xi \geq 0$$

$$\eta_k = C - \alpha_k$$

las siguientes casos son para ξ, o sea para cuando esta arriba de la reg.

Para el caso 1: $\alpha_k = 0 \Rightarrow \eta_k = C \Rightarrow \xi_k = 0$

$$y_k - W^T \phi(x_k) - b \leq \epsilon$$

esta dentro del tubo, entre las dos lineas punteadas y no le aporta al modelo

9 marzo, 22

Caso 2: $0 \leq \alpha_k \leq C$

$y_k - w^T \phi(x_k) - b = \epsilon$

Los vectores estan sobre del tubo punteado

Caso 3: $\alpha_k = C \Rightarrow \eta_k = 0 \Rightarrow \xi_k > 0$

$y_k - w^T \phi(x_k) - b < \epsilon + \xi_k$

esta por afuera del tubo punteado

Finally, the following case considers the application of L_2^{ϵ}-SVR.

Problem 2: The L_2^{ϵ} SVR

Consider the following optimization problem:

$$\begin{aligned}\min_{w,b,\zeta,\zeta^*} \mathcal{P}_c\left(w,b,\zeta,\zeta^*\right) =& \frac{1}{2}w^T w + \frac{c}{2}\sum_{k=1}^{N}\left(\zeta_k^2 + \zeta_k^{*2}\right)\\ \text{s. t. } & y_k - w^T\varphi\left(x_k\right) - b \leq \epsilon + \zeta_k,\ k = 1,\ldots,N\\ & w^T\varphi\left(x_k\right) + b - y_k \leq \epsilon + \zeta_k^*,\ k = 1,\ldots,N\end{aligned} \tag{2}$$

where $\varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ and the regularization parameter $c > 0$ determines the balance between the regularity of f and the quantity up to which we tolerate deviations more significant than ϵ. Consider ζ_k and ζ_k^* as slack variables that control the error between the prediction $\hat{y}_k$ and the k-th sample y_k.
For this case, present a complete development of the problem (2), by using the KKT optimality approach.

16 marzo, 22

problema 2 : ~~L_2 SVR~~ L2 soft-margin SVR

En el problema 1 utilizamos la función de perdida L1.

Ahora vamos a usar L2 que indica errores menores al cuadrado que se relacionan a la norma 1 y norma 2

A esta formulación de SVR se le conoce como squared epsilon insensitivo, la zona insensitiva es epsilon dentro de las lineas punteadas

primero definimos lagrange del dual como

$$\mathcal{L}(w, b, \xi_k, \xi_k^*, \alpha^*, \alpha, \cancel{A}, \cancel{A^*})$$ siendo $\alpha, \alpha^*, \cancel{A}, \cancel{A^*}$ los valores de lagrange

$$= \frac{1}{2} w^T w + \frac{C}{2} \sum (\xi_k^2 + \xi_k^{*2}) - \sum \alpha_k (\epsilon + \xi_k - y_k + w^T \phi(x_k) + b) - \sum \alpha_k^* (\epsilon + \xi_k^* - w^T \phi(x_k) - b + y_k)$$ I

como ξ_k y ξ_k^* estan al cuadrado en el primal, no estan restringidos a la condición $\xi_k \geq 0$, $\xi_k^* \geq 0$. Por lo tanto eliminamos los multiplicadores de lagrange asociados.

Ahora derivamos e igualamos a cero.

$$\frac{\partial L}{\partial w} = \nabla w = w - \sum (\alpha_k - \alpha_k^*) \phi(x_k) = 0 \quad \text{II}$$

$$w = \sum (\alpha_k - \alpha_k^*) \phi(x_k)$$

$$\frac{\partial L}{\partial b} = \sum (\alpha_k^* - \alpha_k) = 0 \quad \text{III}$$

$$\frac{\partial L}{\partial \xi_k} = C \xi_k - \alpha_k = 0 \quad \text{IV}$$

$$\frac{\partial L}{\partial \xi_k^*} = C \xi_k^* - \alpha_k^* = 0 \quad \text{V}$$

16 marzo 2020

Sustituyendo y simplificando II, III, IV y V en I obtenemos el dual

$$\max_{\alpha, \alpha^*} -\frac{1}{2}\sum\sum(\alpha_l-\alpha_l^*)(\alpha_k-\alpha_k^*)\,\varphi(x_k)^T\varphi(x_k)$$

$$-\epsilon\sum(\alpha_k+\alpha_k^*)+\sum y_k(\alpha_k-\alpha_k^*) \qquad \text{VI}$$

s.t. $\sum(\alpha_k-\alpha_k^*)=0$

$\alpha_k \geq 0$ para $k=1,\ldots,N$

$\alpha_k^* \geq 0$ para $k=1,\ldots,N$

Más aún podemos definir $K(x_k,x)=\varphi^T(x_k)\varphi(x_k)$

e introducir $F(x)=\sum(\alpha_k-\alpha_k^*)K(x_k,x)+b \qquad \text{VII}$

donde $\sum(\alpha_k-\alpha_k^*)\varphi(x_k)=W$ (11)

Abe define el dual como

$$\frac{1}{2}\sum_{k,l=1}^{N}(\alpha_k-\alpha_k^*)(\alpha_l-\alpha_l^*)\left(K(x_k,x_l)+\frac{\delta_{k,l}}{C}\right)$$

$$-\epsilon\sum_{k=1}^{N}(\alpha_k+\alpha_k^*)+\sum_{k=1}^{N}(\alpha_k-\alpha_k^*)y_k$$

donde $\delta_{k,l}$ es la función delta de Kronecker

La solución óptima debe satisfacer las siguientes condiciones complementarias

KKT4

$\alpha_k(\epsilon+\xi_k-y_k+W^T\varphi(x_k)+b)=0$ para $k=1,\ldots,N$

$\alpha_k^*(\epsilon+\xi_k^*+y_k-W^T\varphi(x_k)-b)=0$ para $k=1,\ldots,N$

$C\xi_k=\alpha_k$, $C\xi_k^*=\alpha_k^*$ para $k=1,\ldots,N$

16 marzo, 22

Caso 1: $\alpha_k = 0$, $\alpha_k^* = 0$

Todo lo de adentro de la función no importa porque esta multiplicado por 0.

Satisfaze $|y - f(x)| < \epsilon$ y estos vectores no contribuyen a la construccion del modelo

Caso 2: $\alpha_k > 0$, $\alpha_k^* > 0$

$b = y_k - w^T \varphi(x_k) - \epsilon - \frac{\alpha_k}{C}$ para $\alpha_k > 0$

$b = y_k - w^T \varphi(x_k) + \epsilon + \frac{\alpha_k^*}{C}$ para $\alpha_k^* > 0$

estas son los vectores de soporte

L2 no tiene vectores de soporte restringidas y la matriz Hessiana es definida positiva