AU311, Pattern Recognition Tutorial (Fall 2022) Homework: 2. Support Vector Machine

2. Support Vector Machine

Lecturer: Xiaolin Huang `xiaolinhuang@sjtu.edu.cn`

Student: XXX `xxx@sjtu.edu.cn`

Problem 1

There have been many variants of SVM for different purpose. The following is called ν-SVM which can controls the ratio of support vectors. The primal formulation of ν-SVM is given below,

$$\begin{aligned} \min_{w,\rho,\xi} \quad & \frac{1}{2}\|w\|_2^2 - \nu\rho + \frac{1}{m}\sum_{i=1}^{m}\xi_i \\ \text{s.t.} \quad & y_i(w^T x_i + b) \geq \rho - \xi_i \\ & \rho \geq 0, \xi_i \geq 0, \forall i = 1, 2, \ldots m. \end{aligned} \tag{1}$$

Please derive its dual problem and discuss the meaning of ν.

Answer:

The derivation of the dual problem is as follows:

$$L(w,\xi,\nu,\rho,\alpha,\beta) = \frac{1}{2}\|w\|_2^2 - \nu \cdot p + \frac{1}{m}\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i\left(y_i\left(w_T \cdot x + b_i\right) - \rho + \xi_i\right) - \sum_{i=1}^{m}\beta_i\xi_i - \rho\gamma = 0. \tag{2}$$

So, we can have the derivation of the above Lagrangian function:

$$\frac{\partial L}{\partial w} = ||w||_2 - \sum_{i=1}^{m}\alpha_i y_i x_i = 0 \tag{3}$$

$$\frac{\partial L}{\partial \rho} = -\nu + \sum \alpha_i - \gamma = 0 \tag{4}$$

$$\frac{\partial L}{\partial \xi_i} = \frac{1}{m} - \alpha_i - \beta_i = 0 \tag{5}$$

$$\frac{\partial L}{\partial b} = \alpha_i y_i = 0 \tag{6}$$

Substitute equation 3-6 into equation 2, we have:

$$L(w,\xi,\nu,\rho,\alpha,\beta) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_j y_j\left(x_i \cdot x_j\right) \tag{7}$$

So ,we can easily know the dual problem is:

$$
\begin{aligned}
\min_{\alpha} \quad & \tfrac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i y_i x_i x_j y_j \alpha_j \\
\text{s.t.} \quad & \sum_{i=1}^{m} \alpha_i y_i = 0 \\
& \sum_{i=1}^{m} \alpha_i \geq \nu \\
& 0 \leq \alpha_i \leq \tfrac{1}{m}, \forall i = 1, 2, \ldots m
\end{aligned} \tag{8}
$$

From the complementary relaxation condition, we can know that when a sample point is a support vector, its corresponding coefficient $\alpha_i = 0$. And because we are bound by the formula (8). To minimize the number of support vectors, we choose

$$
\begin{aligned}
& \sum_{i=1}^{m} \alpha_i = \nu \\
& 0 \leq \alpha_i \leq \tfrac{1}{m}, \forall i = 1, 2, \ldots m
\end{aligned} \tag{9}
$$

So there is $m\nu$ support vectors **at least**. So, ν is used to control the ratio of support vectors.

Problem 2

The support vector machine can also be used for regression, which is called SVR specifically. Suppose the training data are $\{x_i, y_i\}_{i=1}^{m}$ with $x_i \in R^n, y_i \in R$. Then the regressor $f(x) = w^T x + b$ can be trained by the following primal problem:

$$
\begin{aligned}
\min_{w,\xi,\zeta} \quad & \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{m}(\xi_i + \zeta_i) \\
\text{s.t.} \quad & y_i - w^T x_i - b \leq \varepsilon + \xi_i \\
& w^T x_i + b - y_i \leq \varepsilon + \zeta_i, \\
& \xi_i \geq 0, \zeta_i \geq 0, \forall i = 1, 2, \ldots m.
\end{aligned} \tag{10}
$$

i) explain the loss function used in SVR; *Hint. give the plot of loss v.s. residual.*

ii) derive the dual problem and explain which samples are the support vectors.

Answer:

i)We defined the residual $\mathbb{R}$ is :

$$\mathbb{R} = |y_i - w^T x_i - b| \tag{11}$$

the penalty function is defined as(assuming that $\varepsilon \geq 0$):

$$L(u) = \max\{0, \mathbb{R} - \varepsilon\} \tag{12}$$

so the loss function can be written as:

$$\frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{m} L\left(\left|y_i - w^{\mathrm{T}} x_i - b\right|\right) \tag{13}$$

Then, the plot of loss and residual is below:

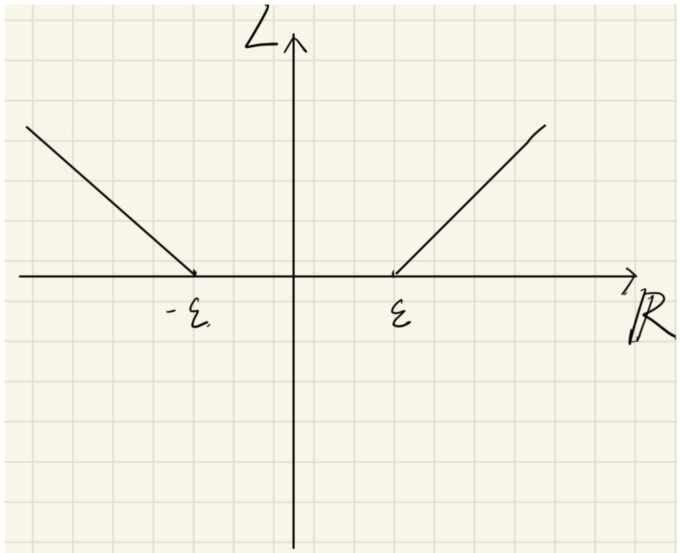

Figure 1: the relation of loss and residual

ii)

The derivation of the dual problem is as folllows:

$$L(w,b,\xi,\zeta)=\frac{1}{2}\|\omega\|_2^2+C\sum_{i=1}^{m}(\xi_i+\zeta_i)+\sum_{i=1}^{m}\alpha_i\left(y_i-\omega^T x_i-b-(\varepsilon+\xi_i)\right)+\sum_{i=1}^{m}\beta_i\left(\omega^T x_i+b-y_i-(\varepsilon+\zeta_i)\right)+\lambda\xi_i+\gamma\zeta_i. \tag{14}$$

So, we can have the derivation of the above Lagrangian function:

$$\frac{\partial L}{\partial w}=\|w\|_2-\sum_{i=1}^{m}\alpha_i x_i+\sum_{i=1}^{m}\beta_i x_i=0 \tag{15}$$

$$\frac{\partial L}{\partial \xi_i}=C-\alpha_i+\lambda=0 \tag{16}$$

$$\frac{\partial L}{\partial \zeta_i}=C-\beta_i+\gamma=0 \tag{17}$$

$$\frac{\partial L}{\partial b}=-\alpha_i+\beta_i=0 \tag{18}$$

Substituting Equation 14-17 into Equation 13, we can get:

$$L(w,b,\xi,\zeta)=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i-\beta_i)\,x_i^{\mathrm{T}}x_j\,(\alpha_j-\beta_j)-\varepsilon\sum_{i=1}^{m}(\alpha_i+\beta_i) \tag{19}$$

So ,we can easily know the dual problem is:

$$\begin{aligned}
\min_{\alpha,\beta} \quad & \tfrac{1}{2} \textstyle\sum_{i=1}^{m} \sum_{j=1}^{m} \left(\alpha_i - \beta_i\right) x_i^{\mathrm{T}} x_j \left(\alpha_j - \beta_j\right) - \varepsilon \sum_{i=1}^{m} \left(\alpha_i + \beta_i\right) \\
\text{s.t.} \quad & \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \beta_i = 0 \\
& 0 \leq \alpha_j \leq C \\
& 0 \leq \beta_j \leq C, \forall j = 1, 2, \ldots m.
\end{aligned} \tag{20}$$

According to the complementary relaxation condition, we can know the support vectors should meet the follow conditions:

$$0 < \alpha_i \leq C \quad \text{and} \quad \beta_i = 0 \tag{21}$$

or

$$0 < \beta_i \leq C \quad \text{and} \quad \alpha_i = 0 \tag{22}$$