HW5 Answer

MLTA

November 2022

Problem 1 (Kernel)

Find the circle with centre in $(0, 0)$ and the radius of 4

Problem 2 (SVM with Gaussian kernel)

1. For a training example (x_i, y_i), we get

$$\begin{aligned}
|f(x_i) - y_i| &= \left|\sum_{j=1}^{m} y_j K(x_j, x_i) - y_i\right| \\
&= \left|\sum_{j=1}^{m} y_j \exp\left(-\|x_j - x_i\|^2 / \tau^2\right) - y_i\right| \\
&= \left|y_i + \sum_{j \neq i} y_j \exp\left(\|x_j - x_i\|^2 / \tau^2\right) - y_i\right| \\
&= \left|\sum_{j \neq i} y_j \exp\left(-\|x_j - x_i\|^2 / \tau^2\right)\right| \\
&\leq \sum_{j \neq i} \left|y_j \exp\left(-\|x_j - x_i\|^2 / \tau^2\right)\right| \\
&= \sum_{j \neq i} |y_j| \cdot \exp\left(-\|x_j - x_i\|^2 / \tau^2\right) \\
&= \sum_{j \neq i} \exp\left(-\|x_j - x_i\|^2 / \tau^2\right) \\
&\leq \sum_{j \neq i} \exp\left(-\epsilon^2 / \tau^2\right) \\
&= (m-1) \exp\left(-\epsilon^2 / \tau^2\right).
\end{aligned} \tag{1}$$

Th, we need to choose a τ such that

$$\tau < \frac{\epsilon}{\log(m-1)}$$

By choosing, for example, $\tau = \epsilon / \log m$ we are done.

Problem 3 (Support Vector Regression)

1. Let $\alpha_i, \alpha_i^*, \beta \geq 0 (i = 1, \cdots, m)$ be the Lagrange multiplier for the primal problem. Then the Lagrangian can be written as:

$$\begin{aligned}
&L\left(w, b, \xi, \alpha, \alpha^*, \beta,\right) \\
&= \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\beta_i\xi_i \\
&- \sum_{i=1}^{m}\alpha_i\left(\epsilon + \xi_i - y_i + w^T x_i + b\right) \\
&- \sum_{i=1}^{m}\alpha_i^*\left(\epsilon + \xi_i + y_i - w^T x_i - b\right)
\end{aligned} \tag{2}$$

2. Note that by $\alpha_i^{(*)}$, we refer to α_i and α_i^*. First, the dual function can be written as:

$$\theta(\alpha, \alpha^*, \beta) = \inf_{w,b,\xi} L\left(w, b, \xi, \alpha, \alpha^*, \beta\right) \tag{3}$$

Now, taking the derivatives of Lagrangian w.r.t. all primal variables, we get

$$\frac{\partial}{\partial w}L = w - \sum_{i=1}^{m}\left(\alpha_i - \alpha_i^*\right)x_i = 0 \Rightarrow w = \sum_{i=1}^{m}\left(\alpha_i - \alpha_i^*\right)x_i \tag{4}$$

$$\frac{\partial}{\partial b}L = \sum_{i=1}^{m}\left(\alpha_i^* - \alpha_i\right) = 0 \tag{5}$$

$$\frac{\partial}{\partial \xi}L = C - \alpha_i^{(*)} - \beta_i = 0 \tag{6}$$

Note that

$$\begin{aligned}
\theta_D\left(\alpha, \alpha^*, \beta\right) = \frac{1}{2}\|w\|^2 - \epsilon\sum_{i=1}^{m}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{m}y_i\left(\alpha_i - \alpha_i^*\right) + b\sum_{i=1}^{m}\left(\alpha_i^* - \alpha_i\right) \\
+ \sum_{i=1}^{m}\left(\alpha_i^* - \alpha_i\right)w^T x_i + \sum_{i=1}^{m}(C - \beta_i - \alpha_i - \alpha_i^*)\xi_i
\end{aligned} \tag{7}$$

By the above equation(4)(5) and (6), we get

$$
\begin{aligned}
\theta_D(\alpha,\alpha^*) &= \frac{1}{2}\|w\|^2 - \epsilon\sum_{i=1}^m(\alpha_i+\alpha_i^*) + \sum_{i=1}^m y_i(\alpha_i-\alpha_i^*) + \sum_{i=1}^m(\alpha_i^*-\alpha_i)\,w^T x_i \\
&= \frac{1}{2}\left\|\sum_{i=1}^m(\alpha_i-\alpha_i^*)\,x_i\right\|^2 - \sum_{i=1}^m(\alpha_i-\alpha_i^*)\left(\sum_{j=1}^m(\alpha_j-\alpha_j^*)\,x_j^T x_i\right) \\
&\quad - \epsilon\sum_{i=1}^m(\alpha_i+\alpha_i^*) + \sum_{i=1}^m y_i(\alpha_i-\alpha_i^*) \\
&= -\frac{1}{2}\sum_{i=1,j=1}^m(\alpha_i-\alpha_i^*)\left(\alpha_j-\alpha_j^*\right)x_i^T x_j - \epsilon\sum_{i=1}^m(\alpha_i+\alpha_i^*) + \sum_{i=1}^m y_i(\alpha_i-\alpha_i^*)
\end{aligned}
\tag{8}
$$

Now, the dual problem can be formulated as:

$$
\begin{aligned}
\max_{\alpha_i,\alpha_i^*} \quad & -\frac{1}{2}\sum_{i=1,j=1}^m(\alpha_i-\alpha_i^*)\left(\alpha_j-\alpha_j^*\right)x_i^T x_j - \epsilon\sum_{i=1}^m(\alpha_i+\alpha_i^*) + \sum_{i=1}^m y_i(\alpha_i-\alpha_i^*) \\
\text{s.t.} \quad & \sum_{i=1}^m(\alpha_i^*-\alpha_i) = 0 \\
& 0 \le \alpha_i, \alpha_i^* \le C
\end{aligned}
\tag{9}
$$

3. (a) Write the primal problem in the form:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^m \max\left(|y_i - \left(w^T x_i + b\right)| - \epsilon, 0\right)$$

Since $\bar{w}$ is optimal, the optimal bias $\bar{b}$ is

$$\underset{b}{\operatorname{argmin}} \sum_{i=1}^m \max\left(\left|y_i - \bar{w}^T x_i + b\right| - \varepsilon, 0\right)$$

(b) Since $(\bar{w}, \bar{b}, \bar{\xi})$ and $(\bar{\alpha}, \bar{\alpha^*}, \bar{\beta})$ satisfies the KKT conditions that the following satisfies for all $i = 1, \cdots, N$:

$(S_1)\, \sum_{i=1}^m(\bar{\alpha}_i - \bar{\alpha}_i{}^*) = 0$	$(P_1)\, y_i - \left(\overline{w}^T x_i + \bar{b}\right) - \epsilon - \bar{\xi}_i \le 0$
$(S_2)\, C = \bar{\alpha}_i + \bar{\alpha}_i{}^* + \bar{\beta}_i$	$(P_2)\, \left(\overline{w}^T x_i + \bar{b}\right) - y_i - \epsilon - \bar{\xi}_i \le 0$
$(S_3)\, \bar{w} = \sum_{i=1}^m(\bar{\alpha}_i - \bar{\alpha}_i^*)\, x_i$	$(P_3)\, -\bar{\xi}_i \le 0$
$(D_1)\, \bar{\alpha}_i, \bar{\alpha}_i^*, \bar{\beta}_i \ge 0$	$(C_1)\, \bar{\alpha}_i\left(y_i - \left(\overline{w}^T x_i + \bar{b}\right) - \epsilon - \bar{\xi}_i\right) = 0$
	$(C_2)\, \bar{\alpha}_i^*\left(\left(\overline{w}^T x_i + \bar{b}\right) - y_i - \epsilon - \bar{\xi}_i\right) = 0$
	$(C_3)\, \bar{\beta}_i\left(-\bar{\xi}_i\right) = 0$

(C_3) is rewritten as $(C - \bar{\alpha}_i - \bar{\alpha}_i^*)\xi_i = 0$ by (S_2)
Define $e = y_i - \left(\overline{w}^T x_i + \bar{b}\right)$

- If $|e| < \epsilon$, then $\bar{\alpha}_i = \bar{\alpha}_i^* = 0$ by $(C_1)(C_2)$, $\bar{\xi}_i = 0$ by (C_3)
- If $e = \epsilon$, then $\bar{\alpha}_i^* = 0$ by (C_2), $\bar{\xi}_i = 0$, $0 \leq \bar{\alpha}_i \leq C$ by $(C_1)(C_3)(D_1)$
- If $e = -\epsilon$, then $\bar{\alpha}_i = 0$ by (C_1), $\bar{\xi}_i = 0$, $0 \leq \bar{\alpha}_i^* \leq C$ by $(C_2)(C_3)(D_1)$
- If $e > \epsilon$, then $\bar{\alpha}_i^* = 0$ by (C_2), $\bar{\xi}_i \neq 0$ by (P_1), $\bar{\alpha}_i = C$ by (C_3), $\bar{\xi}_i = e - \epsilon$ by (C_1)
- If $e < -\epsilon$, then $\bar{\alpha}_i = 0$ by (C_1), $\bar{\xi}_i \neq 0$ by (P_2), $\bar{\alpha}_i^* = C$ by (C_3), $\bar{\xi}_i = -e - \epsilon$ by (C_2)

In fact, $\bar{\alpha}_i \bar{\alpha}_i{}^* = 0$ (its easily to prove by contradiction)

4. By equation (10) in (b), we have $w = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) x_i$, then

$$f(w, x) = w^T x + b = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) x_i^T x + b = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (10)$$

This shows that the decision function can be written as a kernel form.