Homework Set 4, CPSC 8420, Spring 2022

Song, Zhiyuan

April 15, 2022

Problem 1

Considering soft margin SVM, where we have the objective and constraints as follows:

$$\begin{aligned} &min \ \ \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{m}\xi_i \\ &s.t. \ \ y_i(w^T x_i{+}b) \geq 1-\xi_i \ \ (i=1,2,...m) \\ &\xi_i \geq 0 \ \ (i=1,2,...m) \end{aligned} \tag{1}$$

Now we formulate another formulation as:

$$\begin{aligned} &min \ \ \frac{1}{2}||w||_2^2 + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2 \\ &s.t. \ \ y_i(w^T x_i{+}b) \geq 1-\xi_i \ \ (i=1,2,...m) \end{aligned} \tag{2}$$

1. In the Eq.1, we will always have $\exists\theta_i \in \mathbb{R}$ such that $2\theta_i^2 = \xi_i$ because $\xi_i \geq 0$. Meanwhile, $1-\xi_i = 1-2\theta_i^2 \ \ s.t. \ \ \theta_i \in \mathbb{R}$ has the same range with $1-\theta_i \ \ s.t. \ \ \theta \geq 0$. Therefore, Eq.1 can be written as

$$\begin{aligned} &min \ \ \frac{1}{2}||w||_2^2 + \frac{C}{2}\sum_{i=1}^{m}\theta_i^2 \\ &s.t. \ \ y_i(w^T x_i{+}b) \geq 1-\theta_i \ \ (i=1,2,...m) \end{aligned} \tag{3}$$

   If we replace θ_i in Eq.3 with ξ_i, we will end up with the same expression with Eq.2. Accordingly, the ptimal value of the objective will be the same when ξ_i constrain is removed in Eq.2.

2. According to the objective function, we can obtain the generalized Lagrangian of the new soft margin SVM optimization problem.

$$L(w,b,\xi_i) = \frac{1}{2}\|w\|_2^2 + \frac{C}{2}\sum_{i}^{m}\xi_i^2 - \sum_{i}^{m}\alpha_i\left[y_i\left(w^T x_i + b\right) - 1 + \xi_i\right] \tag{4}$$

3. Minimization of $L(w, b, \xi_i)$ leads to partial derivatives with respect to the corresponding variables.

$$\begin{cases} \dfrac{\partial L}{\partial w} & = w - \displaystyle\sum_i^m \alpha_i y_i x_i = 0 \\ \dfrac{\partial L}{\partial b} & = \displaystyle\sum_i^m \alpha_i y_i = 0 \\ \dfrac{\partial L}{\partial \xi_i} & = C\xi_i - \alpha_i = 0 \end{cases}$$

4. We can plug the minimization results from the problem above to get the dual of the version soft margin SVM optimization.

$$\begin{aligned} & max \quad \frac{1}{2}\sum_i^m \sum_j^m \alpha_i \alpha_j y_i y_j < x_i, x_j > + \frac{C}{2}\sum_i^m \xi_i^2 - \sum_i^m \alpha_i \left[y_i \left(\sum_j^m \alpha_j y_j x_j^T x_i + b \right) - 1 + \xi_i \right] \\ = & max \quad -\frac{1}{2}\sum_i^m \sum_j^m \alpha_i \alpha_j y_i y_j < x_i, x_j > + \frac{C}{2}\sum_i^m \xi_i^2 - \sum_i^m \alpha_i y_i b + \sum_i^m \alpha_i - \sum_i^m \alpha_i \xi_i \\ = & min \quad \frac{1}{2}\sum_i^m \sum_j^m \alpha_i \alpha_j y_i y_j < x_i, x_j > + \sum_i^m \left(\alpha_i \xi_i - \frac{C}{2}\xi_i^2 \right) - \sum_i^m \alpha_i \\ = & min \quad \frac{1}{2}\sum_i^m \sum_j^m \alpha_i \alpha_j y_i y_j < x_i, x_j > + \frac{1}{2C}\sum_i^m \alpha_i^2 - \sum_i^m \alpha_i \end{aligned} \tag{5}$$

5. Small C indicates a small penalty affecting optimization in Eq.4. The slack variables are not required to be very small for the minimization process. Therefore, more tolerant misclassifications will be performed. As C increases, the penalty starts to play an important role in Eq.4. The slack variables will be smaller. The classification will be more and more strict.

Problem 2

Recall vanilla SVM objective:

$$L(w, b, \alpha) = \frac{1}{2}||w||_2^2 - \sum_{i=1}^m \alpha_i [y_i(w^T x_i + b) - 1] \quad s.t. \quad \alpha_i \geq 0 \tag{6}$$

If we denote the margin as γ, and vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_m]$, now please show $\gamma^2 * \|\alpha\|_1 = 1$.