

# Homework Set 4, CPSC 8420, Spring 2022

Last Name, First Name

**Due 04/17/2022, Sunday, 11:59PM EST**

## Problem 1

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

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

Now we formulate another formulation as:

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

1. Different from Eq. (1), we now drop the non-negative constraint for  $\xi_i$ , please show that optimal value of the objective will be the same when  $\xi_i$  constraint is removed.

In eqn. (2), since the new regulational term  $\frac{C}{2} \xi_i^2$  is positive by definition, so there is no need to impose the positive constraints on  $\xi_i$ .

2. What's the generalized Lagrangian of the new soft margin SVM optimization problem?

The generalized Lagrangian objective function is:

$$\mathbf{L}(w, b, \alpha, \beta) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i [y_i(w^T x_i + b) - 1 + \xi_i] \tag{3}$$

3. Now please minimize the Lagrangian with respect to  $w, b$ , and  $\xi$ .

In the Lagrangian formulation, we have:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \\ \frac{\partial L}{\partial b} = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \end{cases} \tag{4}$$

, from which we have:

$$\begin{cases} w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \\ \sum_{i=1}^n \alpha_i y_i = 0 \\ C\xi_i - \alpha_i = 0 \end{cases}$$

4. What is the dual of this version soft margin SVM optimization problem? (should be similar to Eq. (10) in the slides)

Plug the results Eqn. (4) from the previous problem into Eqn. (3), we have:

$$\begin{aligned} L(\alpha) &= \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i [y_i (w^T x_i + b) - 1 + \xi_i] \\ &= \frac{1}{2} w^T \sum_{i=1}^n \alpha_i y_i x_i + \frac{C}{2} \sum_{i=1}^n \xi_i^2 - w^T \sum_{i=1}^n \alpha_i y_i x_i - \sum_{i=1}^n \alpha_i y_i b + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i \\ &= -\frac{1}{2} w^T \sum_{i=1}^n \alpha_i y_i x_i + \frac{C}{2} \sum_{i=1}^n \xi_i^2 + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \alpha_i \xi_i \\ &= -\frac{1}{2} \left( \sum_{i=1}^n \alpha_i y_i x_i \right)^T \sum_{i=1}^n \alpha_i y_i x_i + \sum_{i=1}^n \alpha_i + \frac{C}{2} \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i \xi_i \\ &= -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^n \alpha_i + \frac{1}{2} \frac{1}{C} \sum_{i=1}^n \alpha_i^2 - \frac{1}{C} \sum_{i=1}^n \alpha_i^2 \\ &= -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^n \alpha_i - \frac{1}{2C} \sum_{i=1}^n \alpha_i^2 \end{aligned} \tag{5}$$

5. Please analysis bias-variance trade-off when  $C$  increases.

$\xi_i$  is the distance of the falsely classified data point  $x_i$  to its corresponding margin. Therefore,  $C$  in the objective  $L$  weights the errors of the model coming from these non-separable data points. As  $C$  increases, these falsely classified data points will yield more penalty on the optimization objective, and will restrict the margin  $\gamma$  from increasing and including too many unseparable data points. And since  $\gamma = \frac{2}{\|w\|}$ , larger  $C$  leads to larger  $\|w\|$ . To conclude, when  $C$  decreases the bias and increases the variance.

## 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  $\gamma$ , and vector  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_m]$ , now please show  $\gamma^2 * \|\alpha\|_1 = 1$ .

In vanilla SVM, to minimize the objective  $L(w, b, \alpha)$ , we have  $\frac{\partial L}{\partial w} = w - \sum_{i=1}^m \alpha_i y_i x_i = 0$ , from which we get

$$w = \sum_{i=1}^m \alpha_i y_i x_i \quad (7)$$

From the geometric relations, we know that  $\gamma = \frac{1}{\|w\|_2}$ . Therefore, the equation to be proven  $\gamma^2 \|\alpha\|_1 = 1$  is equivalent to:

$$\begin{aligned} \|\alpha\|_1 &= \sum_{i=1}^m |\alpha_i| \\ &= \sum_{i=1}^m \alpha_i \\ &= 1/\gamma^2 \\ &= \|w\|_2^2 \\ &= \left( \sum_{i=1}^m \alpha_i y_i x_i \right)^2 \\ &= \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j \end{aligned} \quad (8)$$

, where the absolute values on  $\alpha_i$  were dropped since we have the constraint  $\alpha_i \geq 0$  by definition. From eqn. (8), equivalent to the original problem, we can alternatively prove:  $\sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^m \alpha_i = 0$ . From  $\frac{\partial L}{\partial w} = 0$ , we have  $\sum_{i=1}^m \alpha_i y_i = 0$ . Plug into eqn. (8), we have:

$$\begin{aligned}
& \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^m \alpha_i \\
&= \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^m \alpha_i + b \times \sum_{i=1}^m \alpha_i y_i \\
&= \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j + b \sum_{i=1}^m \alpha_i y_i - \sum_{i=1}^m \alpha_i \\
&= \sum_{i=1}^m \alpha_i y_i x_i^T \sum_{j=1}^m \alpha_j y_j x_j + b \sum_{i=1}^m \alpha_i y_i - \sum_{i=1}^m \alpha_i \\
&= \sum_{j=1}^m \left( \sum_{i=1}^m \alpha_i y_i x_i^T \right) \alpha_j y_j x_j + b \sum_{i=1}^m \alpha_i y_i - \sum_{i=1}^m \alpha_i \\
&= \sum_{j=1}^m w^T \alpha_j y_j x_j + b \sum_{i=1}^m \alpha_i y_i - \sum_{i=1}^m \alpha_i \\
&= \sum_{i=1}^m w^T \alpha_i y_i x_i + b \sum_{i=1}^m \alpha_i y_i - \sum_{i=1}^m \alpha_i \\
&= \sum_{i=1}^m \alpha_i y_i (w^T x_i) + \sum_{i=1}^m \alpha_i y_i b - \sum_{i=1}^m \alpha_i \\
&= \sum_{i=1}^m \alpha_i [y_i (w^T x_i + b) - 1]
\end{aligned} \tag{9}$$

, which is the same form as the Lagrangian term in the objective function  $L(w, b, \alpha) = \frac{1}{2} \|w\|_2^2 - \sum_{i=1}^m \alpha_i [y_i (w^T x_i + b) - 1]$ . For this term, we always have  $\sum_{i=1}^m \alpha_i [y_i (w^T x_i + b) - 1] = 0$ . Combining eqn. (7)-(9), we have:  $\gamma^2 * \|\alpha\|_1 = 1$ .