

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 ξ_i , please show that optimal value of the objective will be the same when ξ_i constraint is removed.

The optimal value will still be the same since the objective is still convex and therefore be differentiated. Question 1.3 shows that the Lagrangian of this formulation can still be minimized and 1.4 shows that the dual version is the same as the dual for equation 1.

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

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

3. Now please minimize the Lagrangian with respect to w, b , and ξ .

To minimize the Lagrangian we take the derivative with respect to w, b, ξ and set to 0.

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^m \alpha_i y_i x_i = 0 \tag{4}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^m \alpha_i y_i = 0 \tag{5}$$

$$\frac{\partial L}{\partial \xi} = C\xi_i - \alpha_i = 0 \quad (6)$$

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

To derive the dual we apply the variables above to the Lagrangian.

$$\begin{aligned} \underbrace{\min}_{w,b,\xi} L(w,b,\xi,\alpha) &= \frac{1}{2}\|w\|^2 + \frac{C}{2} \sum_{i=1}^m \xi_i^2 - \sum_{i=1}^m \alpha_i [y_i(w^T x_i, b) - 1 + \xi_i] \\ \underbrace{\min}_{w,b,\xi} L(w,b,\xi,\alpha) &= \frac{1}{2}\|w\|^2 - \sum_{i=1}^m \alpha_i [y_i(w^T x_i, b) - 1] \\ &= \frac{1}{2} \left\| \sum_{i=1}^m \alpha_i y_i x_i \right\|^2 - \sum_{i=1}^m \alpha_i [y_i(w^T x_i, b) - 1] \\ &= \frac{1}{2} w^T \sum_{i=1}^m \alpha_i y_i x_i - w^T \sum_{i=1}^m \alpha_i y_i x_i - \sum_{i=1}^m \alpha_i y_i b + \sum_{i=1}^m \alpha_i \\ &= -\frac{1}{2} w^T \sum_{i=1}^m \alpha_i y_i x_i - \sum_{i=1}^m \alpha_i y_i b + \sum_{i=1}^m \alpha_i \\ &= -\frac{1}{2} \left(\sum_{i=1}^m \alpha_i y_i x_i^T \right) \sum_{i=1}^m \alpha_i y_i x_i - b \sum_{i=1}^m \alpha_i y_i + \sum_{i=1}^m \alpha_i = \\ &= -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^m \alpha_i \\ \underbrace{\max}_{\alpha} -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^m \alpha_i &\text{ which is same as} \\ \underbrace{\min}_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j & \\ \text{s.t. } \sum_{i=1}^m \alpha_i y_i &= 0 \\ \alpha_i \geq 0 & i = 1, 2, \dots, m \end{aligned}$$

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

When C is small, the SVM classifier does not allow a lot of misclassification. The support vector classifier will have low bias but may not generalize well and have high variance. Similarly, If C is large, the number of misclassifications allowed has been increased. This classifier would generalize better but will have a higher amount of bias with lower variance.

Problem 2

Recall vanilla SVM objective:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|_2^2 - \sum_{i=1}^m \alpha_i [y_i (w^T x_i + b) - 1] \quad s.t. \quad \alpha_i \geq 0 \quad (7)$$

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

Since support vectors lie on hyperplanes, for any support vector x_i , $w^T x_i + b = y_i$ and therefore b can be obtained by letting $b = y_i - \sum_{j=1}^m \alpha_j y_j (x_j \cdot x_i)$

If we multiply both sides by $\alpha_i y_i$ then we get:

$$\sum_{j=1}^m \alpha_i y_i b = \sum_{j=1}^m \alpha_i y_i^2 - \sum_{j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

We use the fact that $y_i = 1$ and $\gamma = \frac{1}{\|w\|}$ along with $w = \sum_{i=1}^m \alpha_i y_i x_i$

$$\text{Therefore } 0 = \sum_{i=1}^m \alpha_i - \|w\|^2$$

$$\gamma^2 = \frac{1}{\|w\|^2} = \frac{1}{\sum_{i=1}^m \alpha_i} = \frac{1}{\|\alpha\|} \text{ so } \gamma^2 * \|\alpha\| = 1$$