Homework Set 4, CPSC 8420, Fall 2024

Collins, Matthew

Due 11/19/2024, 11:59PM EST

Problem 1

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

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

Now we formulate another formulation as:

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

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.
2. What's the generalized Lagrangian of the new soft margin SVM optimization problem?
3. Now please minimize the Lagrangian with respect to w, b, and ξ.
4. What is the dual of this version soft margin SVM optimization problem? (should be similar to Eq. (10) in the slides)

P1.1, Answer:

Substitutue $\xi_i = 2\theta_i^2$, where $\theta_i \in \mathbb{R}$. By this substitution:

$$\theta_i^2 \geq 0 \implies \xi_i \geq 0.$$

Substituting $\xi_i = 2\theta_i^2$ into the constraints and objective function of Equation (1), we get:

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i = 1 - 2\theta_i^2.$$

The objective function then becomes:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\xi_i = \min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}2\theta_i^2.$$

Rewriting the scaled terms, the objective function is now:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + 2C\sum_{i=1}^{m}\theta_i^2.$$

Let $\tilde{C} = 2C$. Then the problem can be reformulated as:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + \tilde{C}\sum_{i=1}^{m}\theta_i^2, \quad \text{s.t.} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - 2\theta_i^2, \quad i = 1, 2, \ldots, m.$$

This is equivalent to the optimization problem in Equation (2):

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2, \quad \text{s.t.} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, m.$$

Here, the non-negative constraint $\xi_i \geq 0$ in Equation (1) is no longer explicitly needed because $\theta_i^2 \geq 0$ ensures that $\xi_i = 2\theta_i^2 \geq 0$.

P1.2, Answer:

To derive the generalized Lagrangian for this problem, introduce Lagrange multipliers $\alpha_i \geq 0$ for each constraint $i = 1, 2, \ldots, m$. The Lagrangian is constructed by combining the objective function with the constraints weighted by these multipliers.

The objective function of the optimization problem is:

$$\frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2.$$

Constraints:

$$L(\mathbf{w}, b, \xi_i, \alpha_i) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2 - \sum_{i=1}^{m}\alpha_i\Big(y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1 + \xi_i\Big).$$

Expanding the terms:

$$L(\mathbf{w}, b, \xi_i, \alpha_i) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2 - \sum_{i=1}^{m}\alpha_i\big[y_i\mathbf{w}^\top \mathbf{x}_i + y_i b - 1 + \xi_i\big].$$

where $\alpha_i \geq 0$ are the Lagrange multipliers associated with the inequality constraints.

P1.3, Answer:

The partial derivative of L with respect to $\mathbf{w}$ is:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{m}\alpha_i y_i \mathbf{x}_i.$$

Setting $\frac{\partial L}{\partial \mathbf{w}} = 0$, we find:

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i.$$

The partial derivative of L with respect to b is:

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{m} \alpha_i y_i.$$

Setting $\frac{\partial L}{\partial b} = 0$, we obtain the constraint:

$$\sum_{i=1}^{m} \alpha_i y_i = 0.$$

The partial derivative of L with respect to ξ_i is:

$$\frac{\partial L}{\partial \xi_i} = C\xi_i - \alpha_i.$$

Setting $\frac{\partial L}{\partial \xi_i} = 0$, we find:

$$\xi_i = \frac{\alpha_i}{C}.$$

Thus, the minimization of the Lagrangian yields the following conditions:

$$\mathbf{w} - \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i = 0, \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \quad C\xi_i - \alpha_i = 0.$$

P1.4, Answer:

Substituting $\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$ and $\xi_i = \frac{\alpha_i}{C}$ into the Lagrangian, the dual becomes:

$$L(\alpha) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{2}\sum_{i=1}^{n} \xi_i^2 - \sum_{i=1}^{n} \alpha_i \big(y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1 + \xi_i\big).$$

Expanding $\|\mathbf{w}\|^2$:

$$\|\mathbf{w}\|^2 = \left\| \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i \right\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j.$$

Thus, the term $\frac{1}{2}\|\mathbf{w}\|^2$ becomes:

$$\frac{1}{2}\|\mathbf{w}\|^2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j.$$

The penalty term $\frac{C}{2}\sum_{i=1}^{n} \xi_i^2$ simplifies to:

$$\frac{C}{2}\sum_{i=1}^{n}\xi_i^2 = \frac{C}{2}\sum_{i=1}^{n}\left(\frac{\alpha_i}{C}\right)^2 = \frac{1}{2C}\sum_{i=1}^{n}\alpha_i^2.$$

Combining these terms, the dual becomes:

$$L(\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j + \sum_{i=1}^{n}\alpha_i - \frac{1}{2C}\sum_{i=1}^{n}\alpha_i^2.$$

Finally, the dual optimization problem is:

$$\max_{\alpha}\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j - \frac{1}{2C}\sum_{i=1}^{n}\alpha_i^2,$$

subject to:

$$\sum_{i=1}^{n}\alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C.$$