Homework Set 3, CPSC 8420, Fall 2023

Singh, Parampreet

Due 11/17/2023, Friday, 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)

Problem 2

Recall vanilla SVM objective:

$$L(w,b,\alpha) = \frac{1}{2}||w||_2^2 - \sum_{i=1}^{m}\alpha_i[y_i(w^Tx_i+b)-1] \quad s.t. \quad \alpha_i \geq 0 \tag{3}$$

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

Problem 1

Given :

Soft margin SVM :

$$\min \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{m} \xi_i \quad \text{--- (1)}$$

s.t $y_i(w^T x_i + b) \geq 1 - \xi_i \ (i = 1,2,3 \ldots m)$

$\xi_i \geq 0 \ (i = 1,2,3, \ldots m)$

and an another formulation :

$$\min \frac{1}{2}\|w\|_2^2 + \frac{C}{2}\sum_{i=1}^{m} \xi_i^2 \quad \text{--- (2)}$$

s.t $y_i(w^T x + b) \geq 1 - \xi_i \ (i = 1,2,3 \ldots m)$

Intuition :

In equation 2, the penality term penalizes the square of slack variables ξ_i. If it were to take negative values, the squared term will still be positive. The minimilization of the objective function would naturally drive ξ_i towards zero or positive values. Negative ξ_i means a more correctly classified data point, which does not make sense in case of SVM.

Let's denote equation α as J

$$J = \frac{1}{2} \|w\|_2^2 + \frac{C}{2} \sum_{i=1}^{m} \xi_{ei}^2$$

As we can see, J is a strictly convex function with one global minima where $\frac{\partial J}{\partial \xi_e} = 0$.

Let's ~~comp~~ compute $\frac{\partial J}{\partial \xi_{ei}}$,

$$\frac{\partial J}{\partial \xi_{ei}} = C \xi_{ei}$$

For any $\xi_{ei} < 0$, the gradient is negative, which means that we can minimize J by increasing $\xi_{ei} < 0$ towards zero.

Therefore, even without the $\xi_{ei} \geq 0$ constraint, the optimal value for equation α must be non-negative i.e all $\xi_{ei} < 0$ will be $\xi_{ei} = 0$.

Therefore, the optimal value of objective function is same for both the equations.

Part 2 we only have one constraint, so the Langrangian for the new problem is

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

where $\alpha_i \geq 0$

Part 3

If we denote,

$$J(w, b, \xi) = \frac{1}{2}\|w\|_2^2 + \frac{C}{2}\sum_{i=1}^{m} \xi_i^2$$

then we have

$$J = \max_{\alpha_i \geq 0} L(w, b, \xi, \alpha)$$

Therefore, ~~we~~ for optimization, we now have to minimize J as,

$$\min J(w, b, \xi) = \min_{w, b, \xi} \max_{\alpha_i \geq 0} L(w, b, \xi, \alpha)$$

Now, as $J(w, b, \xi)$ is a convex function we can make use of KKT condition therefore, using KKT condition, $J(w, b, \xi)$ is now equivalent to

$$\min J(w,b,\xi) = \max_{\alpha_i \geq 0} \min_{w,b,\xi} L(w,b,\xi)$$

Minimizing Lagrangian w.r.t w, b, ξ:

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

$$\boxed{w = \sum_{i=1}^{m} \alpha_i y_i x_i}$$

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

$$\frac{\partial L}{\partial \xi_i} = C \cancel{\sum_{i=1}} \xi - \cancel{\sum} \alpha_i \Rightarrow \cancel{\sum_{i=1}} (C\xi_i - \alpha_i) = 0$$

$$\Rightarrow \boxed{C \cancel{\sum_{i=1}} \xi = \cancel{\sum^{m}} \alpha} \Rightarrow \boxed{C\xi = \alpha}$$

Part 4

For dual version, substituting value of w in equation in Part 2:

$$\Rightarrow \frac{1}{2} \sum_{i=1}^{m} \alpha_i y_i x_i \sum_{i=1}^{j} \alpha_i y_i x_i + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2 -$$

$$\sum_{i=1}^{m} \alpha_i [y_i (w^T x_i + b) - 1 + \xi]$$

$$\Rightarrow -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j x_i^T x_j + \frac{C}{2}\sum_{i=1}^{m} \xi_i^2 - \cancel{b\sum_{i=1}^{m} y_i \alpha_i}$$

$$+ \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i \xi_i$$

$$\max_{\alpha_i \geq 0} \left[-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j x_i^T x_j + \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m} \alpha_i \xi_i \right]$$

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

$$\alpha_i \geq 0$$

Problem 2

$$L(w,b,\alpha) = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{m}\alpha_i[y_i(w^T x_i + b) - 1] \quad \text{— (1)}$$

$$\text{s.t. } \alpha_i \geq 0$$

To prove $\gamma^2 \|\alpha\|_1 = 1$

differentiating lagrangian w.r.t w

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

$$\boxed{w_j = \sum_{i=1}^{m} y_i \alpha_i x_i}$$

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

$$\Rightarrow \boxed{\sum_{i=1}^{m}\alpha_i y_i = 0} \quad \text{— (2)}$$

substituting w in (1)

$$\frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{m}\alpha_i[y_i(w^T x_i + b)] + \sum_{i=1}^{m}\alpha_i$$

$$= \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^{m}\alpha_i\left[y_i\left(\sum_{j=1}^{m}\alpha_j y_j x_j\right) + b\right] + \sum_{i=1}^{m}\alpha_i + b\cancel{\sum_{i=1}^{m}\alpha_i y_i}^{\,0}$$

$$= \frac{1}{2}\|w\|_2^2 - \|w\|_2^2 + \sum_{i=1}^{m}\alpha_i$$

$$L(w,b,\alpha) = -\frac{\|w\|^2}{2} + \sum_{i=1}^{m}\alpha_i$$

At optimality, i.e w^*, α^*, b^*,

$$L(w^*, b^*, \alpha^*) = \frac{\|w\|^2}{2} \quad (\text{Primal Value})$$

$$L(w^*, b^*, \alpha^*) = -\frac{\|w\|^2}{2} + \sum_{i=1}^{m} \alpha_i \quad (\text{Dual Value})$$

Equating both

$$\frac{\|w\|^2}{2} = -\frac{\|w\|^2}{2} + \sum_{i=1}^{m} \alpha_i$$

$$\boxed{\|w\|^2 = \sum_{i=1}^{m} \alpha_i}$$

$$\because \|\alpha_i\|_1 = \sum_{i=1}^{m} \alpha_i$$

$$\|w\|^2 = \|\alpha_i\| \quad \text{——} \quad (3)$$

As we know, margin $\gamma = \frac{1}{\|w\|}$

Therefore,

$$\frac{1}{\gamma^2} = \|\alpha_i\|_1$$

$$\boxed{\gamma^2 \|\alpha_i\|_1 = 1}$$