Homework Set 3, CPSC 8420, Fall 2023

Singh, Charanjit

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

CPSC 8420

HW3

Problem 1

(1) Soft margin eqⁿ:

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

$$\text{s.t. } y_i(w^T x_i + b) \geq 1 - \xi_i \ (i=1,2,\dots,m) \quad \text{— (1)}$$

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

Another formulation:

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

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

In eqⁿ (2), the penalty includes the square of slack variable $\xi_i^2 \geq 0 \ \forall \xi$, but in context of SVM, a negative slack variable does not make sense.

Consider eqⁿ (2):

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

J is a strictly convex function with minima at a point where $\frac{\partial J}{\partial \xi} = 0$

i.e. $\frac{\partial J}{\partial \xi_i} = C\xi_i$

For any $\xi_i < 0$, the gradient is negative which means that we can minimize J by increasing $\xi_i < 0$ towards 0.

Therefore, even without $\xi_i \geq 0$ constraint, optimal value for eqⁿ 2 must be non negative i.e. for $\xi_i < 0$ will be $\xi_i = 0$

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

② Generalized Lagrangian of new soft margin:

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

where α is lagrange's multiplier.

③ Minimizing Lagrangian w.r.t. w, b & ξ:

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

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

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

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

$$\Rightarrow C\cancel{\sum}\xi = \cancel{\sum}\alpha_i$$

④ For dual version, substituting value of w in equation ①:

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

$$\sum_{i=1}^{m}\alpha_i\left[y_i(w^Tx_i+b)-1+\xi\right)$$

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

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

$$\Rightarrow -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j x_i^T x_j + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2 - b\underbrace{\cancel{\sum_{i=1}^{m} y_i}\alpha_i}_{(=0)} + \sum_{i=1}^{m}\alpha_i$$

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

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

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

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

Problem 2:

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

s.t. $\alpha_i \geq 0$

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

differentiating lagrangian w.r.t. w:

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

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

Now, differentiating w.r.t b:

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

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

Substituting value of w in (1):

we get:

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

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

($=0$, proved above in eqn 2)

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

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

at optimality, ie w^*, α^*, b^*, $L(w,b,\alpha) = \frac{\|w\|^2}{2}$.

[as $L(w,b,\alpha)$ is the lagrangian of $\min \frac{\|w\|^2}{2}$, st. $y_i(w^T x_i + b) \geq 1$]

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

as $\alpha_i \geq 0$, $\|\alpha\|_1 = \sum_{i=1}^{m} \alpha_i$

and if margin $= \gamma$, we know that margin of SVM $= \frac{1}{\|w^*\|} = \gamma$.

so, $$\|w^*\|^2 = \|\alpha_i\|_1$$

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

$$\Rightarrow \boxed{\gamma^2 . \|\alpha\|_1 = 1}$$