SMO algorithm

Tateyama Kaoru

February 9, 2025

In SVM,our optimization problem(with Lagragian duality) is

$$\max_{\boldsymbol{\alpha}} -\frac{1}{2}\sum_{r=1}^{m}\sum_{s=1}^{m}\alpha_r\alpha_s y^{(r)}y^{(s)}K(x^{(r)},x^{(s)})+\sum_{t=1}^{m}\alpha_t$$

with constraints

$$0 \le \alpha_i \le C$$
$$\sum_{i=1}^{m}\alpha_i y^{(i)} = 0$$

where C is the regularization parameter,m is the amount of datas.By the spirit of SMO,we're about using coordinate ascent.Suppose we optimize the objective with respect to α_i and α_j,by the constraint,we have

$$\alpha_i y^{(i)} + \alpha_j y^{(j)} = \gamma$$

where γ is some constant(independent of α_i and α_j).Set

$$W(\alpha_i,\alpha_j) = \sum_{t=1}^{m}\alpha_t - \frac{1}{2}\sum_{r=1}^{m}\sum_{s=1}^{m}\alpha_r\alpha_s y^{(r)}y^{(s)}K(x^{(r)},x^{(s)})$$

We extract the parts that are independent of α_i and α_j,then we have(for the sake of convenience,we let $K(x^{(i)},x^{(j)}) = K_{ij}$)

$$\begin{aligned}W(\alpha_i,\alpha_j) =& \alpha_i + \alpha_j - \frac{1}{2}K_{ii}\alpha_i^2 - \frac{1}{2}K_{jj}\alpha_j^2 - K_{ij}y^{(i)}y^{(j)}\alpha_i\alpha_j \\ & - y^{(i)}\alpha_i\sum_{k\neq i,j}\alpha_k y^{(k)}K_{ik} - -y^{(j)}\alpha_j\sum_{k\neq i,j}\alpha_k y^{(k)}K_{jk}\end{aligned}$$

Now we set $\frac{\partial W}{\partial \alpha_j} = 0$ to get optimization.To cancel α_i,we utilize $\alpha_i = y^{(i)}(\gamma - \alpha_j y^{(j)})$ and thus

$$\begin{aligned}\frac{\partial W}{\partial \alpha_j} =& -(K_{ii}+K_{jj}-2K_{ij})\alpha_j + \gamma y^{(j)}K_{ii} - \gamma y^{(j)}K_{ij} \\ & + y^{(j)}(\sum_{k\neq i,j}\alpha_k y^{(k)}K_{ik} - \sum_{k\neq i,j}\alpha_k y^{(k)}K_{jk}) + 1 - y^{(i)}y^{(j)} \\ =& 0\end{aligned}$$

The sum in the expression is nasty,we need to rephrase it as some quantity that is easier to calculate.To achieve that,we consider the decision function

$$f(x) = \boldsymbol{w}^T\boldsymbol{x} + b = \sum_{r=1}^{m} \alpha_r y^{(r)} K(x^{(i)}, x) + b$$

(recall that we have $\boldsymbol{w} = \sum_{r=1}^{m} \alpha_r y^{(r)} x^{(r)}$ without kernal function.)so we have

$$\begin{aligned}\sum_{k \neq i,j} \alpha_k y^{(k)} K_{ik} =& f(x^{(i)}) - \alpha_i y^{(i)} K_{ii} - \alpha_j y^{(j)} K_{ij} - b \\ \sum_{k \neq i,j} \alpha_k y^{(k)} K_{jk} =& f(x^{(j)}) - \alpha_i y^{(i)} K_{ji} - \alpha_j y^{(j)} K_{jj} - b\end{aligned}$$

Notice that these two equations actually include α_i and α_j,one must realize that this are the parameters not yet updated,so instead,we'd rather write them as

$$\begin{aligned}\sum_{k \neq i,j} \alpha_k y^{(k)} K_{ik} =& f(x^{(i)}) - \alpha_i^{old} y^{(i)} K_{ii} - \alpha_j^{old} y^{(j)} K_{ij} - b \\ \sum_{k \neq i,j} \alpha_k y^{(k)} K_{jk} =& f(x^{(j)}) - \alpha_i^{old} y^{(i)} K_{ji} - \alpha_j^{old} y^{(j)} K_{jj} - b\end{aligned}$$

to tell them apart from the updated α_i and α_j we're seeking for.Now substitute them into the derivative,we obtain

$$\begin{aligned}\frac{\partial W}{\partial \alpha_j} =& y^{(j)}(f(x^{(i)}) - f(x^{(j)})) - (K_{ii} + K_{jj} - 2K_{ij})\alpha_j \\ & (K_{ii} + K_{jj} - 2K_{ij})\alpha_j^{old} + 1 - y^{(i)} y^{(j)} \\ & = 0\end{aligned}$$

Now we introduce the notation $\eta = K_{ii} + K_{jj} - 2K_{ij}$ and $E^{(i)} = f(x^{(i)}) - y^{(i)}$,$E^{(j)} = f(x^{(j)}) - y^{(j)}$ and we have

$$\alpha_j = \alpha_j^{old} + \frac{y^{(j)}(E^{(i)} - E^{(j)})}{\eta}$$

which is the update rule for parameter α_j.However,we have constraint on α_j thus the updated α_j should be

$$\alpha_j^{new} = clip(\alpha_j, L, H) = \begin{cases} H & if \quad \alpha_j > H \\ \alpha_j & if \quad L < \alpha_j < H \\ L & if \quad \alpha_j < L \end{cases}$$

where L and H are lower bound and upper bound of α_j respectively.

After updating α_j,we can immediately update α_i using $\alpha_i y^{((i))} + \alpha_j y^{(j)} = \gamma$ as we have

$$\alpha_i^{new} y^{((i))} + \alpha_j^{new} y^{(j)} = \alpha_i^{old} y^{((i))} + \alpha_j^{old} y^{(j)} = \gamma$$

Therefore,we have the update rule for α_i

$$\alpha_i^{new} = \alpha_i^{old} + y^{(i)} y^{(j)} (\alpha_j^{old} - \alpha_j^{new})$$

Now we'll go to update rule for intersection b.If α_i is support vector(i.e. $L < \alpha_i < H$),by KKT condiction,we must have

$$y^{(i)} (\boldsymbol{w}^T x^{(i)} + b) = 1$$

so we have

$$b = y^{(i)} - \sum_{k=1}^{m} \alpha_k y^{(k)} K_{ik}$$

To calculate updated b,of course we should use updated α_i and α_j,so

$$b^{new} = y^{(i)} - \alpha_i^{new} y^{(i)} K_{ii} - \alpha_J^{new} y^{(i)} K_{ij} - \sum_{k \neq i,j} \alpha_k^{old} y^{(k)} K_{ik}$$

(Note that α's are not updated yet except α_i and α_j)Same as before,we have nasty sum here,still we use the trick to write

$$\sum_{k \neq i,j} \alpha_k^{old} y^{(k)} K_{ik} = f(x^{(i)}) - \alpha_i^{old} y^{(i)} K_{ii} - \alpha_j^{old} y^{(j)} K_{ij} + b^{old}$$

and thus

$$b_1^{new} = b^{old} - E^{(i)} - y^{(i)} K_{ii} (\alpha_i^{new} - \alpha_i^{old}) - y^{(j)} K_{ij} (\alpha_j^{new} - \alpha_j^{old})$$

which is the update rule for b.It is totallt the same case if α_j is support vector where we have

$$b_2^{new} = b^{old} - E^{(j)} - y^{(j)} K_{jj} (\alpha_j^{new} - \alpha_j^{old}) - y^{(i)} K_{ij} (\alpha_i^{new} - \alpha_i^{old})$$

If α_i and α_j are both support vector,one can check that these two b^{new} are equal without efforts.If only one of them is support vector,then we choose the corresponding b^{new} as the updated intersection.If both α_i and α_j are not support vector,instead,we choose $b^{new} = \frac{b_1^{new} + b_2^{new}}{2}$ as the new intersection.