CS 229 PS4 Q1

Chwan-Hao TUNG
861052182
10/29/2016
CS229 PS4 Q1

part a. We want to change a_i and a_j while maintaining $\sum_i \alpha_i y_i = 0$. Therefore,

$$\begin{aligned} \alpha_i y_i + \alpha_j y_j &= \alpha_i' y_i + \alpha_j' y_j \\ \alpha_j' &= \frac{\alpha_i y_i + \alpha_j y_j - \alpha_i' y_i}{y_j} \end{aligned}$$

where α_i' and α_j' are the new values of α_i and α_j respectively.

For the constraint $C \geq \alpha_{j'} \geq 0$, we need to put it in terms of a_i'

$$\begin{array}{rcl} C \geq & \alpha_j' & \geq 0 \\ C \geq & \frac{\alpha_i y_i + \alpha_j y_j - \alpha_i' y_i}{y_j} & \geq 0 \\ C y_j \geq & \alpha_i y_i + \alpha_j y_j - \alpha_i' y_i & \geq 0 \\ \alpha_i + \dfrac{-C y_j + \alpha_j y_j}{y_i} \leq & \alpha_i' & \leq \alpha_i + \dfrac{\alpha_j y_j}{y_i} \end{array}$$

To combine it with the constraint of $0 \leq a_i' \leq C$, for the lower bound we need to take the largest of both lower bounds, and for the upper bounds we need to take the smallest of both upper bound, so we end up with,

$$\max(\alpha_i + \frac{-C y_j + \alpha_j y_j}{y_i}, 0) \leq \quad \alpha_i' \quad \leq \min(\alpha_i + \frac{\alpha_j y_j}{y_i}, C)$$

Need to note that depending on y_i and y_j's sign, the direction of inequality changes.

part b.

By applying the division of G and α as described in part b we get,

$$\begin{aligned} -\frac{1}{2}\alpha^{\intercal} G \alpha + \alpha^{\intercal} 1 &= -\frac{1}{2}\begin{bmatrix} a \ \tilde{a} \end{bmatrix} \begin{bmatrix} H & q^{\intercal} \\ q & \tilde{H} \end{bmatrix} \begin{bmatrix} a \\ \tilde{a} \end{bmatrix} + \begin{bmatrix} a + \tilde{a} \end{bmatrix} \\ &= -\frac{1}{2}(a^{\intercal} \cdot H \cdot a + \tilde{a}^{\intercal} \cdot q \cdot a + a^{\intercal} \cdot q^{\intercal} \cdot \tilde{a} + \tilde{a}^{\intercal} \cdot \tilde{H} \cdot \tilde{a}) + \begin{bmatrix} a + \tilde{a} \end{bmatrix} \end{aligned}$$

Since we don't care about other indices other than i, j when we derive the objective with respect α_i' we can throw away some terms and expand the multiplications,

$$-\frac{1}{2}(a^\intercal \cdot H \cdot a + \tilde{a}^\intercal \cdot q \cdot a + a^\intercal \cdot q^\intercal \cdot \tilde{a}) + a = -\frac{1}{2}(a^\intercal \cdot \begin{bmatrix} K_{ii}y_iy_i & K_{ij}y_iy_j \\ K_{ji}y_jy_i & K_{jj}y_jy_j \end{bmatrix} \cdot a + \tilde{a}^\intercal \cdot q \cdot a + a^\intercal \cdot q^\intercal \cdot \tilde{a}) + a$$
$$= -\frac{1}{2}(\alpha_i'^2 K_{ii}y_i^2 + \alpha_i'\alpha_j' K_{ji}y_jy_i + \alpha_i'\alpha_j' K_{ij}y_iy_j + \alpha_j'^2 K_{jj}y_j^2 + \tilde{a}^\intercal \cdot q \cdot a + a^\intercal \cdot q^\intercal \cdot \tilde{a}) + a$$
$$= -\frac{1}{2}(\alpha_i'^2 K_{ii}y_i^2 + \alpha_i'(\alpha_i y_i + \alpha_j y_j - \alpha_i' y_i)K_{ji}y_i + \alpha_i'(\alpha_i y_i + \alpha_j y_j - \alpha_i' y_i)K_{ij}y_i$$
$$+(\alpha_i y_i + \alpha_j y_j - \alpha_i' y_i)^2 K_{jj} + \tilde{a}^\intercal \cdot q \cdot a + a^\intercal \cdot q^\intercal \cdot \tilde{a}) + a$$

Then we take derivative of a_i' and set it 0 and to get the formula for the extremum of the objective with respect to a_i', and also since y is either 1 or -1, y^2 is just 1.

$$-\alpha_i'(K_{ii} - K_{ji} - K_{ij} + K_{jj}) - \frac{1}{2}[(\alpha_i + \alpha_j y_i y_j)(K_{ji} + K_{ij} - 2K_{jj})$$
$$+(\tilde{a}^\intercal \cdot q \cdot \begin{bmatrix} 1 \\ -\frac{y_i}{y_j} \end{bmatrix} + \begin{bmatrix} 1 & -\frac{y_i}{y_j} \end{bmatrix} \cdot q^\intercal \cdot \tilde{a})] + 1 - \frac{y_i}{y_j} = 0$$

part c.

To find the optimal value for the pair α_i' and α_j' we first take the formula from part b and solve for a_j',

$$a_i' = \frac{-\frac{1}{2}[(\alpha_i + \alpha_j y_i y_j)(K_{ji} + K_{ij} - 2K_{jj}) + (\tilde{a}^\intercal \cdot q \cdot \begin{bmatrix} 1 \\ -\frac{y_i}{y_j} \end{bmatrix} + \begin{bmatrix} 1 & -\frac{y_i}{y_j} \end{bmatrix} q^\intercal \tilde{a})] + 1 - \frac{y_i}{y_j}}{(K_{ii} - K_{ji} - K_{ij} + K_{jj})}$$

We notice how a_i' resolves to a single constant value. This would be the optimal value for the objective function if there were no constraints. But since there are constraints we need to check if this a_i' value is within the feasible region. If a_i' ends up being larger than the upper bound, then we set the upper bound value to be the optimal value of a_i' that maximizes the objective function. If a_i' ends up being less than the lower bound, then we set the lower bound value to be the optimal value of a_i' that maximizes the objective function.