- Nikhil Kamthe
- 861245635
- 10/29/2016
- CS 229
- PS 4

Solution(part a): Let α_i' and α_j' be the new values of α_i and α_j respectively. In order to satisfy the constraint $\sum_i \alpha_i y_i = 0$, we have to keep the sum of $\alpha_i y_i + \alpha_j y_j$ same as before. Which leads us to the following equation. Note that value of y_i can be either -1 or +1.

$$\begin{aligned}
\alpha_i' y_i + \alpha_j' y_j &= \alpha_i y_i + \alpha_j y_j \\
\alpha_j' y_j &= \alpha_i y_i + \alpha_j y_j - \alpha_i' y_i \\
\alpha_j' y_j y_j &= \alpha_i y_i y_j + \alpha_j y_j y_j - \alpha_i' y_i y_j \\
\alpha_j' &= \alpha_i y_i y_j + \alpha_j - \alpha_i' y_i y_j && \text{as } y_j y_j = 1 \\
\alpha_j' &= \alpha_j + y_i y_j(\alpha_i - \alpha_i') \\
\alpha_j' &= \alpha_j + h(\alpha_i - \alpha_i') && \text{where } h = y_i y_j
\end{aligned}$$

Let L be the lower bound and U be the upper bound for α_i'

Case(I): $h = 1$,
$\therefore \alpha_i' + \alpha_j' = k_1$, where k_1 is a constant.

- If $k_1 > C$, then max $\alpha_i = C$ and min $\alpha_i = k_1 - C$
- If $k_1 < C$, then min $\alpha_i = k_1$ and max $\alpha_i = 0$

$$\begin{aligned}
L &= max(0, \alpha_i + \alpha_j - C) \\
U &= min(C, \alpha_i + \alpha_j)
\end{aligned}$$

Case(II): $h = -1$,
$\therefore \alpha_j' - \alpha_i' = k_2$, where k_2 is a constant.

- If $k_2 > 0$, then max $\alpha_i' = 0$, and min $\alpha_i' = C - k_2$.
- If $k_2 < C$, then min $\alpha_i' = -k_2$ and max $\alpha_i' = C$

$$\begin{aligned}
L &= max(0, \alpha_i - \alpha_j) \\
U &= min(C, C + \alpha_i - \alpha_j)
\end{aligned}$$

Thus we now have to restrict α_i within its range i.e. $[L, U]$

- $\alpha_i' = L$, if $\alpha_i < L$
- $\alpha_i' = H$, if $\alpha_i > H$
- $\alpha_i' = \alpha_i'$ (The value which we will compute)

Solution(part b): Let f be the objective function for the dual formulation for support vector machine.

$$f = -\frac{1}{2}\alpha^T G\alpha + \alpha^T 1$$

In order to find the extremum (maximum or minimum) of the objective with respect to α_i, we find the derivative of our objective function with respect to α_i and equate it to zero.

$$\begin{aligned}
\frac{df}{d\alpha_i} &= 0 \\
\frac{d}{d\alpha_i}(-\frac{1}{2}\alpha^T G\alpha + \alpha^T 1) &= 0 \\
\frac{d}{d\alpha_i}(-\frac{1}{2}\alpha^T G\alpha) + \frac{d}{d\alpha_i}\alpha^T 1 &= 0 \\
\frac{da}{d\alpha_i}\frac{d}{da}(-\frac{1}{2}\alpha^T G\alpha) + \frac{da}{d\alpha_i}\alpha^T 1 &= 0 \\
\frac{da}{d\alpha_i}(\frac{d}{da}(-\frac{1}{2}\alpha^T G\alpha)) + \frac{d}{d\alpha_i}(\alpha^T 1) &= 0 \\
\begin{bmatrix}1 & -h\end{bmatrix}(\frac{d}{da}(-\frac{1}{2}(a^T Ha + \tilde{a}^T qa + a^T q^T \tilde{a} + \tilde{a}^T \tilde{H}\tilde{a}))) + (1-h) &= 0 \\
\begin{bmatrix}1 & -h\end{bmatrix}(-\frac{1}{2}(2Ha + q^T\tilde{a} + q^T\tilde{a} + 0)) + (1-h) &= 0 \\
-\begin{bmatrix}1 & -h\end{bmatrix}(Ha + q^T\tilde{a}) + (1-h) &= 0 \qquad \text{...}(\mathbf{i}) \\
\begin{bmatrix}1 & -h\end{bmatrix}(Ha) &= (1-h) - [1-h](q^T\tilde{a}) \\
\begin{bmatrix}1 & -h\end{bmatrix}\begin{bmatrix}G_{ii} & G_{ij} \\ G_{ji} & G_{jj}\end{bmatrix}\begin{bmatrix}\alpha_i' & \alpha_j'\end{bmatrix} &= (1-h) - [1-h](q^T\tilde{a}) \\
\begin{bmatrix}G_{ii} - hG_{ji} & G_{ij} - hG_{jj}\end{bmatrix}\begin{bmatrix}\alpha_i' \\ \alpha_j + s\alpha_i - s\alpha_i'\end{bmatrix} &= (1-h) - [1-h](q^T\tilde{a}) \\
G_{ii}\alpha_i' - hG_{ji}\alpha_i' + G_{ij}\alpha_j + hG_{ij}\alpha_i - sG_{ij}\alpha_i' - sG_{jj}\alpha_j - G_{jj}\alpha_i + G_{jj}\alpha_i' &= (1-h) - [1-h](q^T\tilde{a}) \\
\alpha_i'(G_{ii} - hG_{ji} - hG_{ij} + G_{jj}) + (\alpha_j + h\alpha_i)(G_{ij} - hG_{jj}) &= (1-h) - [1-h](q^T\tilde{a}) \qquad \text{...}(\mathbf{ii})
\end{aligned}$$

$$\boxed{\alpha_i' = \frac{1 - h - [1-h](q^T\tilde{a}) - (\alpha_j + h\alpha_i)(G_{ij} - hG_{jj})}{(G_{ii} - hG_{ji} - hG_{ij} + G_{jj})}}$$

Checking if its a maxima:

$$\begin{aligned}
\frac{d^2 f}{d\alpha_i'^2} &< 0 \\
\frac{d^2}{d\alpha_i'^2}(-\frac{1}{2}\alpha^T G\alpha + \alpha^T 1) &< 0 \\
\frac{d}{d\alpha_i'}(-\begin{bmatrix}1 & -h\end{bmatrix}(Ha + q^T\tilde{a}) + (1-h)) &< 0 \quad \text{using } (i) \\
\frac{d}{d\alpha_i'} - (\alpha_i'(G_{ii} - hG_{ji} - hG_{ij} + G_{jj}) + (\alpha_j + h\alpha_i)(G_{ij} - hG_{jj})) + (1-h) - [1-h](q^T\tilde{a}) &< 0 \quad \text{using } (ii) \\
G_{ii} - hG_{ji} - hG_{ij} + G_{jj} &> 0
\end{aligned}$$

Solution(part c):

Algorithm 1 Local optimization General SVM

procedure UPDATE(i,j)
    $h \leftarrow y_i y_j$;
    $[L, U] \leftarrow$ FETCHBOUNDS();
    if CHECKIFMAXIMA() **then** ▷ Using second derivative to check if its a maxima
        $\alpha_i' =$ COMPUTEALPHA();
        if $\alpha_i' < L$ **then**
            $\alpha_i' \leftarrow L$;
        else
            $\alpha_i' \leftarrow U$;
    else
        $Lobjective =$ COMPUTEOBJECTIVE(L);
        $Uobjective =$ COMPUTEOBJECTIVE(U);
        if $Lobjective > Uobjective$ **then**
            $\alpha_i' \leftarrow L$;
        else
            $\alpha_i' \leftarrow U$;
    if $abs(\alpha_i - \alpha_i') > 10^{-10}(\alpha_i + \alpha_i' + 10^{-10})$ **then**
        $\alpha_j' \leftarrow \alpha_j + h(\alpha_i - \alpha_i')$;
        $\alpha =$ UPDATEALPHA(α_i');
        $b =$ UPDATEINTERCEPT();
        $nch \leftarrow 1$;
    else
        $nch \leftarrow 0$;