Appendix for SVM

1 Lagrange dual function (pp. 16)

As shown in page 15, calculating the derivatives of the Lagrangian with respect to ω and b respectively gives

$$\omega = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} \tag{1}$$

and

$$\sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \tag{2}$$

We then substitute (1) and (2) into the Lagrangian to get the Lagrange dual function as follows

$$\begin{aligned}
\mathcal{L}(\omega, b, \alpha) &= \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{m} \alpha_i [y^{(i)}(\omega^T x^{(i)} + b) - 1] \\
&= \frac{1}{2}\omega^T\omega - \sum_{i=1}^{m} \alpha_i y^{(i)} \omega^T x^{(i)} - \sum_{i=1}^{m} \alpha_i y^{(i)} b + \sum_{i=1}^{m} \alpha_i \\
&= \frac{1}{2}\omega^T \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} - \omega^T \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} - \sum_{i=1}^{m} \alpha_i y^{(i)} b + \sum_{i=1}^{m} \alpha_i \\
&= -\frac{1}{2}\omega^T \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} - \sum_{i=1}^{m} \alpha_i y^{(i)} b + \sum_{i=1}^{m} \alpha_i \\
&= -\frac{1}{2}\left(\sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}\right)^T \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} - b\sum_{i=1}^{m} \alpha_i y^{(i)} + \sum_{i=1}^{m} \alpha_i \\
&= -\frac{1}{2}\sum_{i=1}^{m} \alpha_i y^{(i)} (x^{(i)})^T \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} - b\sum_{i=1}^{m} \alpha_i y^{(i)} + \sum_{i=1}^{m} \alpha_i \\
&= -\frac{1}{2}\sum_{i=1,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^T x^{(j)} - b\sum_{i=1}^{m} \alpha_i y^{(i)} + \sum_{i=1}^{m} \alpha_i \\
&= \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^T x^{(j)} - b\sum_{i=1}^{m} \alpha_i y^{(i)} \\
&= \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^T x^{(j)}
\end{aligned}$$

Therefore,

$$\begin{aligned}\mathcal{G}(\alpha) &= \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}(x^{(i)})^T x^{(j)} \\ &= \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1,j=1}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)} < x^{(i)}, x^{(j)} > \qquad (3)\end{aligned}$$

2 Corollaries on Page 34

If $\alpha_i = 0$**,** $y^{(i)}(\omega^T x^{(i)} + b) \geq 1$

$$\begin{aligned}&\because \alpha_i = 0, \alpha_i + r_i = C \\ &\therefore r_i = C \\ &\because r_i\xi_i = 0 \\ &\therefore \xi_i = 0 \\ &\because y^{(i)}(\omega^T x^{(i)} + b) + \xi_i - 1 \geq 0 \\ &\therefore y^{(i)}(\omega^T x^{(i)} + b) \geq 1\end{aligned}$$

If $\alpha_i = C$**,** $y^{(i)}(\omega^T x^{(i)} + b) \leq 1$

$$\begin{aligned}&\because \alpha_i = C, \alpha_i + r_i = C \\ &\therefore r_i = 0 \\ &\because r_i\xi_i = 0, \xi \geq 0 \\ &\therefore \xi_i \geq 0 \\ &\because \alpha_i(y^{(i)}(\omega^T x^{(i)} + b) + \xi_i - 1) = 0 \\ &\therefore y^{(i)}(\omega^T x^{(i)} + b) + \xi_i - 1 = 0 \\ &\therefore y^{(i)}(\omega^T x^{(i)} + b) = 1 - \xi \leq 1\end{aligned}$$

If $0 < \alpha_i < C$**,** $y^{(i)}(\omega^T x^{(i)} + b) = 1$

$$\begin{aligned}&\because 0 < \alpha_i < C, \alpha_i + r_i = C \\ &\therefore 0 < r_i < C \\ &\because r_i\xi_i = 0, \xi \geq 0 \\ &\therefore \xi_i = 0 \\ &\because \alpha_i(y^{(i)}(\omega^T x^{(i)} + b) + \xi_i - 1) = 0 \\ &\therefore y^{(i)}(\omega^T x^{(i)} + b) + \xi_i - 1 = 0 \\ &\therefore y^{(i)}(\omega^T x^{(i)} + b) = 1\end{aligned}$$

3 SMO Algorithm

The dual objective function (see pp. 38) can be written as

$$\begin{aligned} f(\alpha_1,\alpha_2) &= \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}K_{ij} \\ &= \alpha_1+\alpha_2+\sum_{i=3}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\alpha_i\alpha_j y^{(i)}y^{(j)}K_{ij} \\ &\quad -\sum_{i=1}^{2}\sum_{j=3}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}K_{ij} - \frac{1}{2}\sum_{i=3}^{m}\sum_{j=3}^{m}\alpha_i\alpha_j y^{(i)}y^{(j)}K_{ij} \end{aligned}$$

Let $V_i = \sum_{j=3}^{m} y^{(j)}\alpha_j K_{ij}$, then

$$\begin{aligned} f(\alpha_1,\alpha_2) &= \alpha_1+\alpha_2-\frac{1}{2}\alpha_1^2K_{11} - y^{(1)}y^{(2)}\alpha_1\alpha_2K_{12} \\ &\quad -\frac{1}{2}\alpha_2^2K_{22} - y^{(1)}\alpha_1V_1 - y^{(2)}\alpha_2V_2 + const1 \end{aligned}$$

where $const1$ is independent with both α_1 and α_2.

Since $\alpha_1 = (\zeta - \alpha_2 y^{(2)})y^{(1)}$ (see pp. 37),

$$\begin{aligned} f(\alpha_2) &= -\frac{1}{2}(K_{11}-2K_{12}+K_{22})\alpha_2^2 \\ &\quad +y^{(2)}(y^{(2)}-y^{(1)}+\zeta K_{11}-\zeta K_{12}+V_1-V_2)\alpha_2 + const2 \end{aligned}$$

As shown in page 39, let the derivative of $f(\alpha_2)$ be zero

$$\begin{aligned} \frac{\partial}{\partial\alpha_2}f(\alpha_2) &= -(K_{11}-2K_{12}+K_{22})\alpha_2 \\ &\quad +y^{(2)}(y^{(2)}-y^{(1)}+\zeta K_{11}-\zeta K_{12}+V_1-V_2) \\ &= 0 \end{aligned}$$

then,

$$(K_{11}-2K_{12}+K_{22})\alpha_2 = y^{(2)}(y^{(2)}-y^{(1)}+\zeta K_{11}-\zeta K_{12}+V_1-V_2)$$

Assuming

$$f_i = \sum_{j=1}^{m} y^{(j)}\alpha_j K_{ij} + b$$

$$E_i = f_i - y^{(i)}$$

$$V_i = \sum_{j=3}^{m} y^{(j)}\alpha_j K_{ij} = \sum_{j=1}^{m} y^{(j)}\alpha_j K_{ij} - \sum_{j=1}^{2} y^{(j)}\alpha_j K_{ij} = f_i - \sum_{j=1}^{2} y^{(j)}\alpha_j K_{ij} - b$$

and α^+ denotes the updated value (while α is the old one), we have

$$\begin{aligned}
(K_{11} - 2K_{12} + K_{22})\alpha_2^+ &= y^{(2)}(y^{(2)} - y^{(1)} + \zeta K_{11} - \zeta K_{12} + V_1 - V_2) \\
&= y^{(2)}[y^{(2)} - y^{(1)} + \zeta(K_{11} - K_{12}) \\
&\quad +(f_1 - y^{(1)}\alpha_1 K_{11} - y^{(2)}\alpha_2 K_{12} - b) \\
&\quad -(f_2 - y^{(1)}\alpha_1 K_{21} - y^{(2)}\alpha_2 K_{22} - b)] \\
&= y^{(2)}[(f_1 - y^{(1)}) - (f_2 - y^{(2)}) + \zeta(K_{11} - K_{12}) \\
&\quad -(\zeta - y^{(2)}\alpha_2)K_{11} - y^{(2)}\alpha_2 K_{12} \\
&\quad +(\zeta - y^{(2)}\alpha_2)K_{12} - y^{(2)}\alpha_2 K_{22}] \\
&= y^{(2)}[(f_1 - y^{(1)}) - (f_2 - y^{(2)}) + y^{(2)}\alpha_2(K_{11} - 2K_{12} + K_{22})] \\
&= (K_{11} - 2K_{12} + K_{22})\alpha_2 + y^{(2)}(E_1 - E_2)
\end{aligned}$$

and thus

$$\alpha_2^+ = \alpha_2 + \frac{y^{(2)}(E_1 - E_2)}{K_{11} - 2K_{12} + K_{22}}$$

We now compute α_1^+. For α_1 and α_2, we have $\alpha_1 + \alpha_2 y^{(1)} y^{(2)} = y^{(1)}\zeta$. Then, for α_1^+ and α_2^+, we also have

$$\alpha_1^+ + \alpha_2^+ y^{(1)} y^{(2)} = y^{(1)}\zeta = \alpha_1 + \alpha_2 y^{(1)} y^{(2)}$$

Therefore, we get

$$\alpha_1^+ = \alpha_1 + (\alpha_2 - \alpha_2^+) y^{(1)} y^{(2)}$$

4 Calculating b (pp. 40)

When $\alpha_1 \in (0, C)$, we have $y^{(1)}(w^T x^{(1)} + b) = 1$. Since $w = \sum_{i=1}^m \alpha_i y^{(i)} x^{(i)}$,

$$\begin{aligned}
& y^{(1)} \left(\sum_{i=1}^m \alpha_i y^{(i)} x^{(i)} \right)^T x^{(1)} + y^{(1)} b \\
= \quad & y^{(1)} \left(\alpha_1 y^{(1)} x^{(1)} + \alpha_2 y^{(2)} x^{(2)} + \sum_{i=3}^m \alpha_i y^{(i)} x^{(i)} \right)^T x^{(1)} + y^{(1)} b \\
= \quad & \alpha_1 K_{11} + y^{(1)} y^{(2)} \alpha_2 K_{12} + \sum_{i=3}^m \alpha_i y^{(1)} y^{(i)} K_{1i} + y^{(1)} b \\
= \quad & 1
\end{aligned}$$

In fact, the above equation can be written as

$$\alpha_1^+ K_{11} + y^{(1)} y^{(2)} \alpha_2^+ K_{12} + \sum_{i=3}^m \alpha_i y^{(1)} y^{(i)} K_{1i} + y^{(1)} b^+ = 1$$

As $y^{(1)}, y^{(2)} \in \{1, -1\}$, we have

$$y^{(1)} \alpha_1^+ K_{11} + y^{(2)} \alpha_2^+ K_{12} + \sum_{i=3}^m \alpha_i y^{(i)} K_{1i} + b^+ = y^{(1)}$$

Recalling that $f_1 = \sum_{i=1}^m \alpha_i y^{(i)} K_{1i} + b$ and $E_1 = f_1 - y^{(1)}$,

$$y^{(1)} = \sum_{i=1}^{m} \alpha_i y^{(i)} K_{1i} + b - E_1$$

Therefore,

$$\begin{aligned}
b^+ &= y^{(1)} - \alpha_1^+ K_{11} - y^{(2)} \alpha_2^+ K_{12} - \sum_{i=3}^{m} \alpha_i y^{(i)} K_{1i} \\
&= \sum_{i=1}^{m} \alpha_i y^{(i)} K_{1i} + b - E_1 - \alpha_1^+ K_{11} - y^{(2)} \alpha_2^+ K_{12} - \sum_{i=3}^{m} \alpha_i y^{(i)} K_{1i} \\
&= (\alpha_1 - \alpha_1^+) y^{(1)} k_{11} + (\alpha_2 - \alpha_2^+) y^{(2)} k_{12} + b - E_1 \\
&= b_1
\end{aligned}$$

Similarly, when $\alpha_2 \in (0, C)$, we have

$$\begin{aligned}
b^+ &= (\alpha_1 - \alpha_1^+) y^{(1)} k_{12} + (\alpha_2 - \alpha_2^+) y^{(2)} k_{22} + b - E_2 \\
&= b_2
\end{aligned}$$

Also when $\alpha_2, \alpha_2 \in (0, C)$, $b^+ = b_1 = b_2$, and when $\alpha_2, \alpha_2 \in \{0, C\}$, b^+ can take any value between b_1 and b_2. One common choice is

$$b^+ = \frac{b_1 + b_2}{2}$$