6.3

$$\left\{\begin{array}{ll}\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i+b \geqslant+1, & y_i=+1 \\ \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i+b \leqslant-1, & y_i=-1\end{array}\right.$$

[推导]：假设这个超平面是$\left(\boldsymbol{w}^{\prime}\right)^{\top} \boldsymbol{x}+b^{\prime}=0$，对于$\left(\boldsymbol{x}_i, y_i\right) \in D$，有：

$$\left\{\begin{array}{ll}\left(\boldsymbol{w}^{\prime}\right)^{\top} \boldsymbol{x}_i+b^{\prime}>0, & y_i=+1 \\ \left(\boldsymbol{w}^{\prime}\right)^{\top} \boldsymbol{x}_i+b^{\prime}<0, & y_i=-1\end{array}\right.$$

根据几何间隔，将以上关系修正为：

$$\left\{\begin{array}{ll}\left(\boldsymbol{w}^{\prime}\right)^{\top} \boldsymbol{x}_i+b^{\prime} \geq+\zeta, & y_i=+1 \\ \left(\boldsymbol{w}^{\prime}\right)^{\top} \boldsymbol{x}_i+b^{\prime} \leq-\zeta, & y_i=-1\end{array}\right.$$

其中ζ为某个大于零的常数，两边同除以ζ，再次修正以上关系为：

$$\left\{\begin{array}{ll}\left(\frac{1}{\zeta} \boldsymbol{w}^{\prime}\right)^{\top} \boldsymbol{x}_i+\frac{b^{\prime}}{\zeta} \geq+1, & y_i=+1 \\ \left(\frac{1}{\zeta} \boldsymbol{w}^{\prime}\right)^{\top} \boldsymbol{x}_i+\frac{b^{\prime}}{\zeta} \leq-1, & y_i=-1\end{array}\right.$$

令：$\boldsymbol{w}=\frac{1}{\zeta} \boldsymbol{w}^{\prime}, b=\frac{b^{\prime}}{\zeta}$，则以上关系可写为：

$$\left\{\begin{array}{ll}\boldsymbol{w}^{\top} \boldsymbol{x}_i+b \geq+1, & y_i=+1 \\ \boldsymbol{w}^{\top} \boldsymbol{x}_i+b \leq-1, & y_i=-1\end{array}\right.$$

6.8

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha})=\frac{1}{2}\|\boldsymbol{w}\|^2+\sum_{i=1}^m \alpha_i\left(1-y_i\left(\boldsymbol{w}^{\top} \boldsymbol{x}_i+b\right)\right)$$

[推导]： 待求目标: $\begin{array}{ll}\min\limits_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ s.t. & h(\boldsymbol{x})=0 \\ & g(\boldsymbol{x}) \leq 0\end{array}$

等式约束和不等式约束：$h(\boldsymbol{x})=0, g(\boldsymbol{x}) \leq 0$分别是由一个等式方程和一个不等式方程组成的方程组。

拉格朗日乘子：$\boldsymbol{\lambda}=\left(\lambda_1, \lambda_2, \ldots, \lambda_m\right) \qquad \boldsymbol{\mu}=\left(\mu_1, \mu_2, \ldots, \mu_n\right)$

拉格朗日函数：$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})=f(\boldsymbol{x})+\boldsymbol{\lambda} h(\boldsymbol{x})+\boldsymbol{\mu} g(\boldsymbol{x})$

6.9-6.10

$$\begin{aligned} w &= \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i \\ 0 &= \sum_{i=1}^m \alpha_i y_i\end{aligned}$$

[推导]：式（6.8）可作如下展开：

$$\begin{aligned} L(\boldsymbol{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2}||\boldsymbol{w}||^2+\sum_{i=1}^m \alpha_i(1-y_i(\boldsymbol{w}^T \boldsymbol{x}_i+b)) \\ &= \frac{1}{2}||\boldsymbol{w}||^2+\sum_{i=1}^m(\alpha_i-\alpha_i y_i \boldsymbol{w}^T \boldsymbol{x}_i-\alpha_i y_i b) \\ &= \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w}+\sum_{i=1}^m \alpha_i-\sum_{i=1}^m \alpha_i y_i \boldsymbol{w}^T \boldsymbol{x}_i-\sum_{i=1}^m \alpha_i y_i b\end{aligned}$$

对$\boldsymbol{w}$和b分别求偏导数并令其等于0：

$\frac{\partial L}{\partial \boldsymbol{w}} = \frac{1}{2} \times 2 \times \boldsymbol{w} + 0 - \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i - 0 = 0 \Longrightarrow \boldsymbol{w} = \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i$

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

6.11

$$\begin{aligned}
\max_{\boldsymbol{\alpha}} & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j \\
s.t. & \sum_{i=1}^m \alpha_i y_i = 0 \\
& \alpha_i \geq 0 \quad i = 1, 2, \ldots, m
\end{aligned}$$

[推导]：将式 (6.9)代人 (6.8) ，即可将$L(\boldsymbol{w}, b, \boldsymbol{\alpha})$ 中的 $\boldsymbol{w}$ 和 b 消去,再考虑式 (6.10) 的约束,就得到式 (6.6) 的对偶问题：

$$\begin{aligned}
\min_{\boldsymbol{w},b} L(\boldsymbol{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \alpha_i y_i \boldsymbol{w}^T \boldsymbol{x}_i - \sum_{i=1}^m \alpha_i y_i b \\
&= \frac{1}{2}\boldsymbol{w}^T \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i - \boldsymbol{w}^T \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i + \sum_{i=1}^m \alpha_i - b \sum_{i=1}^m \alpha_i y_i \\
&= -\frac{1}{2}\boldsymbol{w}^T \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i + \sum_{i=1}^m \alpha_i - b \sum_{i=1}^m \alpha_i y_i
\end{aligned}$$

又$\sum_{i=1}^m \alpha_i y_i = 0$，所以上式最后一项可化为0，于是得：

$$\begin{aligned}
\min_{\boldsymbol{w},b} L(\boldsymbol{w}, b, \boldsymbol{\alpha}) &= -\frac{1}{2}\boldsymbol{w}^T \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i + \sum_{i=1}^m \alpha_i \\
&= -\frac{1}{2}(\sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i)^T (\sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i) + \sum_{i=1}^m \alpha_i \\
&= -\frac{1}{2} \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i^T \sum_{i=1}^m \alpha_i y_i \boldsymbol{x}_i + \sum_{i=1}^m \alpha_i \\
&= \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j
\end{aligned}$$

所以

$\max_{\boldsymbol{\alpha}} \min_{\boldsymbol{w},b} L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \max_{\boldsymbol{\alpha}} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$

6.39

$C = \alpha_i + \mu_i$ [推导]：对式（6.36）关于ξ_i求偏导并令其等于0可得：

$\frac{\partial L}{\partial \xi_i} = 0 + C \times 1 - \alpha_i \times 1 - \mu_i \times 1 = 0 \Longrightarrow C = \alpha_i + \mu_i$

6.40

$$
\begin{aligned}
&\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j \\
&s.t. \sum_{i=1}^{m}\alpha_i y_i=0 \\
&\quad 0\leq\alpha_i\leq C \quad i=1,2,\dots,m
\end{aligned}
$$

将式6.37-6.39代入6.36可以得到6.35的对偶问题：

$$
\begin{aligned}
\min_{\boldsymbol{w},b,\boldsymbol{\xi}} L(\boldsymbol{w},b,\boldsymbol{\alpha},\boldsymbol{\xi},\boldsymbol{\mu}) &= \frac{1}{2}||\boldsymbol{w}||^2+C\sum_{i=1}^{m}\xi_i+\sum_{i=1}^{m}\alpha_i(1-\xi_i-y_i(\boldsymbol{w}^T\boldsymbol{x}_i+b))-\sum_{i=1}^{m}\mu_i\xi_i \\
&= \frac{1}{2}||\boldsymbol{w}||^2+\sum_{i=1}^{m}\alpha_i(1-y_i(\boldsymbol{w}^T\boldsymbol{x}_i+b))+C\sum_{i=1}^{m}\xi_i-\sum_{i=1}^{m}\alpha_i\xi_i-\sum_{i=1}^{m}\mu_i\xi_i \\
&= -\frac{1}{2}\sum_{i=1}^{m}\alpha_i y_i\boldsymbol{x}_i^T\sum_{i=1}^{m}\alpha_i y_i\boldsymbol{x}_i+\sum_{i=1}^{m}\alpha_i+\sum_{i=1}^{m}C\xi_i-\sum_{i=1}^{m}\alpha_i\xi_i-\sum_{i=1}^{m}\mu_i\xi_i \\
&= -\frac{1}{2}\sum_{i=1}^{m}\alpha_i y_i\boldsymbol{x}_i^T\sum_{i=1}^{m}\alpha_i y_i\boldsymbol{x}_i+\sum_{i=1}^{m}\alpha_i+\sum_{i=1}^{m}(C-\alpha_i-\mu_i)\xi_i \\
&= \sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\boldsymbol{x}_i^T\boldsymbol{x}_j
\end{aligned}
$$

所以
$$
\begin{aligned}
\max_{\boldsymbol{\alpha},\boldsymbol{\mu}}\min_{\boldsymbol{w},b,\boldsymbol{\xi}} L(\boldsymbol{w},b,\boldsymbol{\alpha},\boldsymbol{\xi},\boldsymbol{\mu}) &= \max_{\boldsymbol{\alpha},\boldsymbol{\mu}}\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\boldsymbol{x}_i^T\boldsymbol{x}_j \\
&= \max_{\boldsymbol{\alpha}}\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\boldsymbol{x}_i^T\boldsymbol{x}_j
\end{aligned}
$$
又
$$
\begin{aligned}
\alpha_i &\geq 0 \\
\mu_i &\geq 0 \\
C &= \alpha_i+\mu_i
\end{aligned}
$$
消去μ_i可得等价约束条件为：$0\leq\alpha_i\leq C \quad i=1,2,\dots,m$

6.52

$$
\left\{\begin{array}{l}
\alpha_i\left(f\left(\boldsymbol{x}_i\right)-y_i-\epsilon-\xi_i\right)=0 \\
\hat{\alpha}_i\left(y_i-f\left(\boldsymbol{x}_i\right)-\epsilon-\hat{\xi}_i\right)=0 \\
\alpha_i\hat{\alpha}_i=0, \xi_i\hat{\xi}_i=0 \\
\left(C-\alpha_i\right)\xi_i=0,\left(C-\hat{\alpha}_i\right)\hat{\xi}_i=0
\end{array}\right.
$$

[推导]：将式（6.45）的约束条件全部恒等变形为小于等于0的形式可得：

$$
\left\{\begin{array}{l}
f\left(\boldsymbol{x}_i\right)-y_i-\epsilon-\xi_i\leq 0 \\
y_i-f\left(\boldsymbol{x}_i\right)-\epsilon-\hat{\xi}_i\leq 0 \\
-\xi_i\leq 0 \\
-\hat{\xi}_i\leq 0
\end{array}\right.
$$

由于以上四个约束条件的拉格朗日乘子分别为$\alpha_i,\hat{\alpha}_i,\mu_i,\hat{\mu}_i$，所以由西瓜书附录式（B.3）可知，以上四个约束条件可相应转化为以下KKT条件：

$$
\left\{\begin{array}{l}
\alpha_i\left(f\left(\boldsymbol{x}_i\right)-y_i-\epsilon-\xi_i\right)=0 \\
\hat{\alpha}_i\left(y_i-f\left(\boldsymbol{x}_i\right)-\epsilon-\hat{\xi}_i\right)=0 \\
-\mu_i\xi_i=0 \Rightarrow \mu_i\xi_i=0 \\
-\hat{\mu}_i\hat{\xi}_i=0 \Rightarrow \hat{\mu}_i\hat{\xi}_i=0
\end{array}\right.
$$

由式（6.49）和式（6.50）可知：

$$
\begin{aligned}
\mu_i &= C-\alpha_i \\
\hat{\mu}_i &= C-\hat{\alpha}_i
\end{aligned}
$$

所以上述KKT条件可以进一步变形为：

$$\left\{\begin{array}{l}\alpha_i\left(f\left(\boldsymbol{x}_i\right)-y_i-\epsilon-\xi_i\right)=0 \\ \hat{\alpha}_i\left(y_i-f\left(\boldsymbol{x}_i\right)-\epsilon-\hat{\xi}_i\right)=0 \\ \left(C-\alpha_i\right) \xi_i=0 \\ \left(C-\hat{\alpha}_i\right) \hat{\xi}_i=0\end{array}\right.$$

又因为样本$(\boldsymbol{x}_i, y_i)$只可能处在间隔带的某一侧，那么约束条件$f\left(\boldsymbol{x}_i\right)-y_i-\epsilon-\xi_i=0$和$y_i-f\left(\boldsymbol{x}_i\right)-\epsilon-\hat{\xi}_i=0$不可能同时成立，所以$\alpha_i$和$\hat{\alpha}_i$中至少有一个为0，也即$\alpha_i\hat{\alpha}_i=0$。在此基础上再进一步分析可知，如果$\alpha_i=0$的话，那么根据约束$(C-\alpha_i)\xi_i=0$可知此时$\xi_i=0$，同理，如果$\hat{\alpha}_i=0$的话，那么根据约束$(C-\hat{\alpha}_i)\hat{\xi}_i=0$可知此时$\hat{\xi}_i=0$，所以$\xi_i$和$\hat{\xi}_i$中也是至少有一个为0，也即$\xi_i\hat{\xi}_i=0$。将$\alpha_i\hat{\alpha}_i=0, \xi_i\hat{\xi}_i=0$整合进上述KKT条件中即可得到式（6.52）。