(4) $H_{\tilde{w},\tilde{b}}$ correctly classifies the sample pts. In fact,

$$x^{(1)}\cdot\tilde{w}+\tilde{b}\ \overset{\tilde{w}=\frac{2}{b_2-b_1}w}{\underset{\tilde{b}=\frac{b_1+b_2}{b_2-b_1}}{=}}\ \frac{2}{b_2-b_1}\underbrace{x^{(1)}\cdot w}_{=-b_1}+\frac{b_1+b_2}{b_2-b_1}$$

$$=\frac{-2b_1}{b_2-b_1}+\frac{b_1+b_2}{b_2-b_1}=1$$

and if $y_i=1$,

$$x^{(i)}\cdot\tilde{w}+\tilde{b}\ \overset{\tilde{w}=\frac{2}{b_2-b_1}w}{\underset{\tilde{b}=\frac{b_1+b_2}{b_2-b_1}}{=}}\ \frac{2}{b_2-b_1}\underbrace{x^{(i)}\cdot w}_{\geq -b_1 \text{ by the choice of } x^{(1)}}+\frac{b_1+b_2}{b_2-b_1}$$

$$\geq 1$$

$$x^{(2)}\cdot\tilde{w}+\tilde{b}\ \overset{\tilde{w}=\frac{2}{b_2-b_1}w}{\underset{\tilde{b}=\frac{b_1+b_2}{b_2-b_1}}{=}}\ \frac{2}{b_2-b_1}\underbrace{x^{(2)}\cdot w}_{=-b_2}+\frac{b_1+b_2}{b_2-b_1}$$

$$=\frac{-2b_2}{b_2-b_1}+\frac{b_1+b_2}{b_2-b_1}=-1$$

and if $y_j = -1$

$$x^{(j)} \cdot \tilde{w} + \tilde{b} = \frac{2}{b_2 - b_1} \underbrace{x^{(j)} \cdot w}_{\leq -b_2 \text{ by the choice of } x^{(2)}} + \frac{b_1 + b_2}{b_2 - b_1}$$

$$\leq -1.$$

It remains to solve

$$\max_{\tilde{w}, \tilde{b}} \frac{2}{\|\tilde{w}\|} \quad \text{(margin maximization)}$$

$$\text{s.t.} \quad \begin{cases} x^{(i)} \cdot \tilde{w} + \tilde{b} \geq 1 & \text{if } y_i = 1 \\ x^{(i)} \cdot \tilde{w} + \tilde{b} \leq -1 & \text{if } y_i = -1 \end{cases}$$

(correct classification)

This is equivalent to

$$\min_{\tilde{w} \in \mathbb{R}^p, \tilde{b} \in \mathbb{R}} \underbrace{\|\tilde{w}\|^2}_{\stackrel{\text{def}}{=} f(\tilde{w}, \tilde{b})} \qquad (1)$$

$$\text{s.t.} \quad y_i \left(x^{(i)} \cdot \tilde{w} + \tilde{b} \right) \geq 1$$

or $\underbrace{1 - y_i\left(x^{(i)} \cdot \tilde{w} + \tilde{b}\right)}_{h_i(\tilde{w}, \tilde{b})} \leq 0, \quad i = 1, \cdots, n$

Note $f(\tilde{w}, \tilde{b}) = \|\tilde{w}\|^2 = (\tilde{w}, \tilde{b}) \begin{pmatrix} I_{p\times p} & \\ & 0 \end{pmatrix} \begin{pmatrix} \tilde{w} \\ \tilde{b} \end{pmatrix}$

$\mathrm{Hess}(f)(\tilde{w}, \tilde{b}) = 2 \begin{pmatrix} I_{p\times p} & \\ & 0 \end{pmatrix}$ sym, pos. semi-def

Thus $f(\tilde{w}, \tilde{b})$ is a convex function.

$h_i(\tilde{w}, \tilde{b}) = 1 - y_i\left(x^{(i)} \cdot \tilde{w} + \tilde{b}\right)$, as an affine function of $(\tilde{w}, \tilde{b})$, is a convex func.

Thus ① is a convex optimization problem, which can be numerically solved using convex optimization packages.

4.2.3 Algorithm:

Input: (linearly separable) sample pts with labels $(x^{(1)}, y_1), \cdots, (x^{(n)}, y_n)$ where $y_i = \pm 1$

Output: weight $\hat{w} \in \mathbb{R}^p$ and bias $\hat{b} \in \mathbb{R}$ of the separating hyperplane with maximal margin $\{x \in \mathbb{R}^p : x \cdot \hat{w} + \hat{b} = 0\}$

Step 1: solve the convex optimization problem

$$\min_{\hat{w}, \hat{b}} \quad f(\hat{w}, \hat{b}) = \|\hat{w}\|^2$$

$$\text{s.t.} \quad h_i(\hat{w}, \hat{b}) = 1 - y_i \left(x^{(i)} \cdot \hat{w} + \hat{b} \right) \leq 0$$

Reference: "Learning with Kernels: Support

Vector Machine, Regularization, Optimization"
(in the syllabus) by Scholkopf-Smola
Chapter 1.