Bishop 6.2

$$w = \sum_{n=1}^{N} \alpha_n t_n \phi(x_n)$$ $\alpha_n \equiv$ frequency of n updating w

$$y(x) = \text{sign}\left(w^T \phi(x)\right)$$
$$= \text{sign}\left(\sum_{n=1}^{N} \alpha_n t_n \phi(x_n)^T \phi(x)\right)$$
$$= \text{sign}\left(\sum_{n=1}^{N} \alpha_n t_n k(x_n, x)\right)$$

$\therefore\ \alpha_n \rightarrow \alpha_n + 1$

$$t_n\left(w^T \phi(x_n)\right) \geq 0$$

$$t_n\left(\sum_{m=1}^{N} k(x_m, x_n)\right) \geq 0$$

Learning algorithm depends on Gram Matrix

Bishop 7.3

$$w^T x_1 + b = 1$$
$$w^T x_2 + b = -1$$

$$\underset{w,b}{\text{argmin}} \left\{ \frac{1}{2} \|w\|^2 + \underbrace{\lambda}(w^T x_1 + b - 1) + \underbrace{\eta}(w^T x_2 + b + 1) \right\}$$

Lagrange multiplier

Taking the derivative:

$$w + \lambda x_1 + \eta x_2 = 0$$
$$\lambda + \eta = 0$$

$$w = \lambda(x_1 - x_2)$$

$$2b = -w^T(x_1 + x_2)$$
$$b = \frac{-\lambda}{2}(x_1 - x_2)^T(x_1 + x_2)$$
$$= \frac{-\lambda}{2}(x_1^T x_1 - x_2^T x_2)$$

Bishop 7.4

$$\rho = \frac{1}{\|w\|} \rightarrow \frac{1}{\rho^2} = \|w\|^2$$

$$L(w, b, a) = \frac{\|w\|^2}{2} = \sum_{n}^{N} a_n - \frac{\|w\|^2}{2}$$