1.1) $\min_{\vec{w},b} L = -\min_{\vec{w},b} - \sum_n \{ y_n \log[\sigma(\vec{w}^T \vec{x}_n + b)] + (1-y_n)\log[1-\sigma(\vec{w}^T\vec{x}_n + b)] \}$

$z_n = \vec{w}^T \vec{x}_n + b$

$$\frac{\partial L}{\partial z_n} = -\sum_n \left[y_n \left(\frac{1}{\cancel{\sigma(z_n)}\ln 2}\right)\left(\cancel{\sigma(z_n)}(1-\sigma(z_n))\right) + (1-y_n)\left(\frac{1}{\cancel{(1-\sigma(z_n))}\ln 2}\right)\left(-\sigma(z_n)\cancel{(1-\sigma(z_n))}\right)\right]$$

$$= -\sum_n \left[y_n \frac{1-\sigma(z_n)}{\ln 2} - (1-y_n)\frac{\sigma(z_n)}{\ln 2}\right] = -\sum_n \left[y_n(1-\sigma(z_n)) - (1-y_n)\sigma(z_n)\right]$$

$$= -\sum_n \left[y_n - \cancel{y_n\sigma(z_n)} - \sigma(z_n) + \cancel{y_n\sigma(z_n)}\right] = \sum_n \sigma(z_n) - y_n$$

$$\frac{\partial L}{\partial \vec{w}} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial \vec{w}} = \sum_n [(\sigma(z_n) - y_n)(\vec{x}_n)] = \sum_n [\sigma(\vec{w}^T\vec{x}_n + b) - y_n]\vec{x}_n$$

$$= \underline{\underline{X}}^T[\sigma(\underline{\underline{X}}\vec{w} + b) - \vec{y}]$$

$$\vec{w}_{t+1} = \vec{w}_t - \frac{\eta}{N}[\underline{\underline{X}}^T(\sigma(\underline{\underline{X}}\vec{w}_t + b) - \vec{y})]$$

1.2) $w_0 = 0 \quad \eta = 0.001 \quad N=4 \quad \vec{x} = [1\ 1\ 1\ 1]^T \quad \vec{y} = [0\ 1\ 1\ 1]^T$

$$w_1 = w_0 - \frac{\eta}{N}[\vec{x}^T(\sigma(w_0\vec{x}) - \vec{y})] = 0.00025$$

$$\mathbb{I}[\sigma(w_1\vec{x}) \le 0.5] = [1\ 1\ 1\ 1]^T \Rightarrow \frac{3}{4} = 0.75$$

$*\vec{x} = [1\ 1\ 1\ 1]^T$

1.3) $\vec{x} = [-1\ 1\ 1]^T$

$$\mathbb{I}[\sigma(w_1\vec{x}) \le 0.5] = [0\ 1\ 1]^T \Rightarrow \frac{2}{3} = 0.67$$

2.1) Mercer's theorem: k is a kernel iff $\underline{\underline{K}}$ is positive semidefinite for any N and any $\vec{x}_N$

$\underline{\underline{K}} = \underline{\underline{I}} \in \mathbb{R}^{N \times N}$

since $\underline{\underline{K}} = \underline{\underline{I}}$ and $\underline{\underline{I}}$ is positive semidefinite for any N and any $\vec{x}_n$, k is a kernel

2.2) $\vec{W}_{opt} = \underset{\vec{w}}{\operatorname{argmin}} \|\underline{\underline{\Phi}}\vec{w} - \vec{y}\|_2^2 = \underset{\vec{w}}{\operatorname{argmin}} (\underline{\underline{\Phi}}\vec{w} - \vec{y})^T(\underline{\underline{\Phi}}\vec{w} - \vec{y})$

$\nabla_w F = 2\underline{\underline{\Phi}}^T(\underline{\underline{\Phi}}\vec{w} - \vec{y}) = 0$

$\vec{W}_{opt} = \underline{\underline{\Phi}}^T(\underline{\underline{\Phi}}\vec{w} - \vec{y}) \Rightarrow \underline{\underline{\Phi}}^T\vec{\alpha}$ where $\vec{\alpha} = \underline{\underline{\Phi}}\vec{w} - \vec{y}$

$F(\underline{\underline{\Phi}}^T\vec{\alpha}) = \|\underline{\underline{\Phi}}\underline{\underline{\Phi}}^T\vec{\alpha} - \vec{y}\|_2^2 = \|\underline{\underline{K}}\vec{\alpha} - \vec{y}\|_2^2$ since $\underline{\underline{K}} = \underline{\underline{\Phi}}\underline{\underline{\Phi}}^T$

$G(\vec{\alpha}) = F(\underline{\underline{\Phi}}^T\vec{\alpha})$

$\nabla_\alpha G = \cancel{2\underline{\underline{K}}}(\underline{\underline{K}}\vec{\alpha} - \vec{y}) = 0$

$\vec{\alpha}_{opt} = \underline{\underline{K}}^{-1}\vec{y} = \vec{y}$ since $\underline{\underline{K}} = \underline{\underline{I}}$

$F(\underline{\underline{\Phi}}^T\vec{\alpha}_{opt}) = \|\underline{\underline{K}}\vec{\alpha}_{opt} - \vec{y}\|_2^2 = \|\vec{y} - \vec{y}\|_2^2 = 0$

2.3) $\vec{W}_{opt}{}^T\vec{\phi}(\vec{x}) = \sum_{n=1}^{N} \alpha_n k(\vec{x}_n, \vec{x})$

since $\vec{x} \notin \{\vec{x}_1, \ldots, \vec{x}_n\}$, $\vec{x} \neq \vec{x}_n$, $\forall n$

thus $k(\vec{x}_n, \vec{x}) = 0$

$\vec{W}_{opt}^T\vec{\phi}(\vec{x}) = 0$

3.1) <u>primal formulation</u>

$$\min_{\vec{w},b} \frac{1}{2}\|\vec{w}\|_2^2 \quad \text{st.} \quad 1 - y_n(\vec{w}^T\vec{\phi}(\vec{x}_n) + b) \leq 0, \quad \forall n$$

$$\Rightarrow \quad (0,1) \Rightarrow \quad 1 - (\vec{w}^T[1,0]^T + b) \leq 0$$

$$(\tfrac{\pi}{2},-1) \Rightarrow \quad 1 + (\vec{w}^T[0,1]^T + b) \leq 0$$

$$(\pi,1) \Rightarrow \quad 1 - (\vec{w}^T[-1,0]^T + b) \leq 0$$

<u>dual formulation</u>

$$\max_{\{\alpha_n\}} \sum_n \alpha_n - \frac{1}{2}\sum_{m,n} y_m y_n \alpha_n \alpha_m k(\vec{x}_m, \vec{x}_n) \quad \text{st.} \quad \sum_n \alpha_n y_n = 0, \quad 0 \leq \alpha_n, \quad \forall n$$

$$\Rightarrow \max_{\alpha} \left\{ [\alpha_1 + \alpha_2 + \alpha_3] - \frac{1}{2}\left[\overset{1,1}{\alpha_1^2} - \overset{1,2}{\cancel{\alpha_1\alpha_2(0)}} + \overset{1,3}{\alpha_1\alpha_3(-1)} - \overset{2,1}{\cancel{\alpha_2\alpha_1(0)}} + \overset{2,2}{\alpha_2^2(1)} - \overset{2,3}{\cancel{\alpha_2\alpha_3(0)}} + \overset{3,1}{\alpha_3\alpha_1(-1)} + \overset{3,2}{\cancel{\alpha_3\alpha_2(0)}} + \overset{3,3}{\alpha_3^2(1)} \right] \right\}$$

$$\max_{\alpha} \left\{ [\alpha_1 + \alpha_2 + \alpha_3] - \frac{1}{2}[\alpha_1^2 - 2\alpha_1\alpha_3 + \alpha_2^2 + \alpha_3^2] \right\} = \max_{\alpha} -\frac{\alpha_1^2}{2} + \alpha_1 - \frac{\alpha_2^2}{2} + \alpha_2 - \frac{\alpha_3^2}{2} + \alpha_3 + \alpha_1\alpha_3$$

$$\text{st.} \quad \alpha_1 - \alpha_2 + \alpha_3 = 0, \quad 0 \leq \alpha_n, \forall n$$

3.2) $\alpha_1 - \alpha_2 + \alpha_3 = 0$

$$\alpha_2 = \alpha_1 + \alpha_3 \quad \Rightarrow \quad \max_{\alpha} = -\frac{\alpha_1^2}{2} - \frac{\alpha_2^2}{2} - \frac{\alpha_3^2}{2} + 2\alpha_2 + \alpha_1\alpha_3$$

$$\left.\begin{array}{l} \nabla_{\alpha_1} = -\alpha_1 + \alpha_3 = 0 \\ \nabla_{\alpha_2} = -\alpha_2 + 2 = 0 \\ \nabla_{\alpha_3} = -\alpha_3 + \alpha_1 = 0 \end{array}\right\} \quad \begin{array}{l} \alpha_1 = \alpha_3 = 1 \\ \alpha_2 = 2 \end{array}$$

$$\vec{w}_{opt} = \sum_n \alpha_n^* y_n \vec{\phi}(x)$$

$$= [1,0]^T - 2[0,1]^T + [-1,0]^T$$

$$= [0,-2]^T$$

$$\alpha_n^*(1 - y_n(\vec{w}_{opt}^T\vec{\phi}(\vec{x}_n) + b^*)) = 0 \quad \Rightarrow \quad b^* = y_1 - \vec{w}_{opt}^T\vec{\phi}(x) = 1 - [0,-2][1,0]^T$$

$$= 0$$