Problem 4.1

$$J(w) = \frac{1}{2}\sum_{n=1}^{N}\{w^T\phi(x_n) - t_n\}^2 + \frac{\lambda}{2}w^Tw \qquad (6.2)$$

$$\nabla_w J(w) = \frac{1}{2}\sum_{n=1}^{N} 2\{w^T\phi(x_n) - t_n\}\cdot\phi(x_n) \quad + \lambda w = 0$$

$$\Rightarrow w = -\frac{1}{\lambda}\sum_{n=1}^{N}\{w^T\phi(x_n) - t_n\}\cdot\phi(x_n) =$$

$$= \sum_{n=1}^{N}\left(\frac{\{w^T\phi(x_n) - t_n\}}{-\lambda}\right)\cdot\phi(x_n) = \sum_{n=1}^{N=1} a_n\,\phi(x_n) =$$

$$= \Phi^T a \quad (6.3) \qquad a = (a_1, a_2 \dots a_N)^T$$

we define $a_n = \left(-\frac{1}{\lambda}(w^T\phi(x_n) - t_n)\right)$ $\quad (6.4)$

now we substitute $w = \Phi^T a$ into $J(w)$ to obtain

$$J(\Phi^T a) = \frac{1}{2}\sum_{n=1}^{N}\left[(\Phi^T a)^T\cdot\phi(x_n) - t_n\right]^2 + \frac{\lambda}{2}\cdot(\Phi^T a)^T(\Phi^T a)$$

$$= \frac{1}{2}\sum_{n=1}^{N}\left[a^T\Phi\cdot\phi(x_n) - t_n\right]^2 + \frac{\lambda}{2}\cdot a^T\Phi\Phi^T a$$

$$= \frac{1}{2}\sum_{n=1}^{N}\left[a^T\Phi\cdot\phi(x_n) - t_n\right]^T\left[a^T\Phi\cdot\phi(x_n) - t_n\right] \quad + \frac{\lambda}{2}a^T\Phi\Phi^T a$$

$$= \frac{1}{2}a^T\Phi\Phi^T\Phi\Phi^T a - a^T\Phi\Phi^T t + \frac{1}{2}t^T t + \frac{\lambda}{2}a^T\Phi\Phi^T a \quad (6.5)$$

we now define the Gram matrix $\underline{\underline{K}} = \underline{\underline{\Phi}}\,\underline{\underline{\Phi}}^T$

where the (n,m) element is

$$K_{n,m} = \phi(x_n)^T \phi(x_m) = k(x_n, x_m) \quad (6.6)$$

$\Rightarrow$ we can rewrite $J(a)$ as

$$J(a) = \frac{1}{2} a^T (\Phi \Phi^T)(\Phi \Phi^T) a - a^T (\Phi \Phi^T) t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T (\Phi \Phi^T) a$$

$$= \frac{1}{2} a^T K K a - a^T K t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T K a \quad (6.7)$$

we now compute $\nabla_a J(a)$ and set it equal to 0

$\Rightarrow$

$$\nabla_a J(a) = KKa - Kt + \lambda K a = 0$$

$$\Rightarrow \quad KKa + \lambda K a = Kt$$

$$Ka + \lambda a = t$$

$$(K + \lambda I_N) a = t$$

$$a = (K + \lambda I_n)^{-1} t \quad (6.8)$$

We now substitute (6.8) into the linear Regression model given by $y(x) = w^T \phi(x)$

tave $w = \Phi^T a$

$$\Rightarrow \quad y(x) = w^T \phi(x) = a^T \Phi \phi(x) = k(x)^T (K + \lambda I_N)^{-1} t$$

(6.9)

Problem 4.2

a) consider the definition of a kernel given by

$k(x,x') = \phi(x)^T \phi(x')$, where $\phi(x)$ - are a non-linear feature space mapping.

consider $k_1(x,x') = \phi_1(x)^T \phi_1(x')$, a kernel

now consider $k(x,x') = c \cdot k_1(x,x')$, if k is a valid kernel we must be able to find a valid representation using some feature vectors

if we take $\phi(x) = c^{\frac{1}{2}} \phi_1(x)$ we can rewrite

$$\boxed{\begin{aligned} k(x,x') &= c^{\frac{1}{2}} \phi_1(x)^T \cdot c^{\frac{1}{2}} \phi_1(x') = c \cdot \phi_1(x)^T \phi_1(x') \\ &= c \cdot k_1(x,x') \end{aligned}}$$

$\Rightarrow$ $c \cdot k_1(x,x')$ - is a valid kernel

b) Now show that $k(x,x') = f(x) k_1(x,x') f(x')$ is a valid kernel, where $f(x) \to$ any function

we can rewrite $k(x,x')$

$$k(x,x') = f(x) \cdot \phi_1(x)^T \phi_1(x') \cdot f(x')$$

if we define $v(x) = \phi_1(x) \cdot f(x)$, we can write

$$k(x,x') = v(x)^T v(x')$$

$\Rightarrow$ $k(x,x')$ can be written as an scalar product of feature mappings $\Rightarrow$

$k(x,x') = f(x)\, k_1(x,x')\, f(x')$ is a valid kernel

Problem 4.3

suppose $\exists\ w_1, b_1$ that satisfy the solution to equation (7.5) and minimize equation (7.3). Our constraint will have the form

$$t_n(w_1^T \phi(x_n) + b_1) \geq 1$$

then suppose we have a constraint

$$t_n(w^T \phi(x_n) + b) \geq \gamma \quad , \gamma \geq 0$$

if we rescale our variables such that $w = \gamma w_1$ and $b = \gamma b_1$, we will still minimize (7.3) and our constraint would become

$$t_n((\gamma w_1)^T \phi(x_n) + \gamma b_1) \geq \gamma$$

$$\cancel{\gamma} \cdot t_n(w_1^T \phi(x_n) + b_1) \geq \cancel{\gamma}$$

$$t_n(w_1^T \phi(x_n) + b_1) \geq 1$$

therefore replacing 1 on the r.h.s with $\gamma > 0$, does not change the solution for the maximum margin hyperplane