6.10 First we check that the kernel $k(\mathbf{x}, \mathbf{x'}) = f(\mathbf{x})f(\mathbf{x'})$ is a valid kernel.

Based on the solution to exercise 6.5, the Gram matrix corresponding to this kernel is given by :

$$\begin{bmatrix} f(\mathbf{x}_1)^2 & f(\mathbf{x}_1)f(\mathbf{x}_2) & \dots & f(\mathbf{x}_1)f(\mathbf{x}_N) \\ f(\mathbf{x}_2)f(\mathbf{x}_1) & f(\mathbf{x}_2)^2 & \dots & f(\mathbf{x}_2)f(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{x}_N)f(\mathbf{x}_1) & f(\mathbf{x}_N)f(\mathbf{x}_2) & \dots & f(\mathbf{x}_N)^2 \end{bmatrix}$$

$$= \begin{bmatrix} f(\mathbf{x}_1) & 0 & \dots & 0 \\ 0 & f(\mathbf{x}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(\mathbf{x}_N) \end{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} f(\mathbf{x}_1) & 0 & \dots & 0 \\ 0 & f(\mathbf{x}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(\mathbf{x}_N) \end{bmatrix}$$

The matrix of all ones can be represented as an outer product of a vector of ones with itself, a.k.a, $\mathbf{1}\mathbf{1}^T$.

This matrix is positive semidefinite as can be seen:

$$\forall\, \mathbf{x}, \mathbf{x}^T \left(\mathbf{1}\mathbf{1}^T\right) \mathbf{x} = \left(\mathbf{x}^T\mathbf{1}\right)\left(\mathbf{1}^T\mathbf{x}\right) = \left(\mathbf{x}^T\mathbf{1}\right)\left(\mathbf{x}^T\mathbf{1}\right) = \left(\mathbf{x}^T\mathbf{1}\right)^2 \geq 0$$

The Gram matrix will also be PSD (Positive Semidefinite), based on the discussion in exercise 6.5.

The matrix of ones can be decomposed into the following matrix product:

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} = \left(\frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}\right) \left(\frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}\right)$$

Then, the Gram matrix can be written as:

$$\begin{bmatrix} f(\mathbf{x}_1) & 0 & \dots & 0 \\ 0 & f(\mathbf{x}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(\mathbf{x}_N) \end{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} f(\mathbf{x}_1) & 0 & \dots & 0 \\ 0 & f(\mathbf{x}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(\mathbf{x}_N) \end{bmatrix}$$

$$= \begin{bmatrix} f(\mathbf{x}_1) & 0 & \dots & 0 \\ 0 & f(\mathbf{x}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(\mathbf{x}_N) \end{bmatrix} \left(\frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \right) \left(\frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \right) \begin{bmatrix} f(\mathbf{x}_1) & 0 & \dots & 0 \\ 0 & f(\mathbf{x}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(\mathbf{x}_N) \end{bmatrix}$$

$$= \left(\frac{1}{\sqrt{N}} \begin{bmatrix} f(\mathbf{x}_1) & 0 & \dots & 0 \\ 0 & f(\mathbf{x}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(\mathbf{x}_N) \end{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \right) \left(\frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} f(\mathbf{x}_1) & 0 & \dots & 0 \\ 0 & f(\mathbf{x}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(\mathbf{x}_N) \end{bmatrix} \right)$$

$$= \left(\frac{1}{\sqrt{N}} \begin{bmatrix} f(\mathbf{x}_1) & f(\mathbf{x}_1) & \dots & f(\mathbf{x}_1) \\ f(\mathbf{x}_2) & f(\mathbf{x}_2) & \dots & f(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{x}_N) & f(\mathbf{x}_N) & \dots & f(\mathbf{x}_N) \end{bmatrix} \right) \left(\frac{1}{\sqrt{N}} \begin{bmatrix} f(\mathbf{x}_1) & f(\mathbf{x}_2) & \dots & f(\mathbf{x}_N) \\ f(\mathbf{x}_1) & f(\mathbf{x}_2) & \dots & f(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{x}_1) & f(\mathbf{x}_2) & \dots & f(\mathbf{x}_N) \end{bmatrix} \right)$$

$$= \left(\frac{1}{\sqrt{N}} \begin{bmatrix} f(\mathbf{x}_1) & f(\mathbf{x}_1) & \dots & f(\mathbf{x}_1) \\ f(\mathbf{x}_2) & f(\mathbf{x}_2) & \dots & f(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{x}_N) & f(\mathbf{x}_N) & \dots & f(\mathbf{x}_N) \end{bmatrix} \right) \left(\frac{1}{\sqrt{N}} \begin{bmatrix} f(\mathbf{x}_1) & f(\mathbf{x}_1) & \dots & f(\mathbf{x}_1) \\ f(\mathbf{x}_2) & f(\mathbf{x}_2) & \dots & f(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{x}_N) & f(\mathbf{x}_N) & \dots & f(\mathbf{x}_N) \end{bmatrix} \right)^T$$

Let

$$\mathbf{\Phi} = \left(\frac{1}{\sqrt{N}} \begin{bmatrix} f(\mathbf{x}_1) & f(\mathbf{x}_1) & \dots & f(\mathbf{x}_1) \\ f(\mathbf{x}_2) & f(\mathbf{x}_2) & \dots & f(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{x}_N) & f(\mathbf{x}_N) & \dots & f(\mathbf{x}_N) \end{bmatrix} \right)$$

Then, using equation 6.9, the linear model based on this kernel will be:

$$y(\mathbf{x}) = \mathbf{k}(\mathbf{x})^T \left(\mathbf{K} + \lambda \mathbf{I}_N\right)^{-1} \mathbf{t}$$

$$= f(\mathbf{x}) \begin{bmatrix} f(\mathbf{x}_1) & f(\mathbf{x}_2) & \dots & f(\mathbf{x}_N) \end{bmatrix} \left(\mathbf{\Phi}\mathbf{\Phi}^T + \lambda \mathbf{I}_N \right)^{-1} \mathbf{t}$$

$f(\mathbf{x})\,[f(\mathbf{x}_1) \quad f(\mathbf{x}_2) \quad \dots \quad f(\mathbf{x}_N)]$ can be written as:

$$f(\mathbf{x})\,[f(\mathbf{x}_1) \quad f(\mathbf{x}_2) \quad \dots \quad f(\mathbf{x}_N)] = \begin{bmatrix} f(\mathbf{x})f(\mathbf{x}_1) \\ f(\mathbf{x})f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x})f(\mathbf{x}_N) \end{bmatrix}^T$$

$$= \left(\left(\frac{1}{\sqrt{N}} \begin{bmatrix} f(\mathbf{x}_1) & f(\mathbf{x}_1) & \dots & f(\mathbf{x}_1) \\ f(\mathbf{x}_2) & f(\mathbf{x}_2) & \dots & f(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{x}_N) & f(\mathbf{x}_N) & \dots & f(\mathbf{x}_N) \end{bmatrix} \right) \left(\frac{1}{\sqrt{N}} \begin{bmatrix} f(\mathbf{x}) \\ f(\mathbf{x}) \\ \vdots \\ f(\mathbf{x}) \end{bmatrix} \right) \right)^T$$

$$= (\mathbf{\Phi}\boldsymbol{\phi}(\mathbf{x}))^T$$

where $\boldsymbol{\phi}(\mathbf{x}) = \frac{1}{\sqrt{N}} \begin{bmatrix} f(\mathbf{x}) \\ f(\mathbf{x}) \\ \vdots \\ f(\mathbf{x}) \end{bmatrix}$

Thus, the simplified model becomes:

$$y(\mathbf{x}) = (\mathbf{\Phi}\boldsymbol{\phi}(\mathbf{x}))^T \left(\mathbf{K} + \lambda \mathbf{I}_N\right)^{-1} \mathbf{t}$$

$$= (\mathbf{\Phi}\boldsymbol{\phi}(\mathbf{x}))^T \mathbf{a}$$

$$= \mathbf{a}^T \mathbf{\Phi}\boldsymbol{\phi}(\mathbf{x})$$

$$= \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x})$$

$$= \frac{1}{\sqrt{N}} f(\mathbf{x})(\mathbf{w}^T \mathbf{1})$$

proving that the solution is proportional to $f(\mathbf{x})$.