6.22 Let

$$\mathbf{t}_{(1,N)} = \begin{bmatrix} t(\mathbf{x}_1) \\ t(\mathbf{x}_2) \\ \vdots \\ t(\mathbf{x}_N) \end{bmatrix}, \quad \mathbf{t}_{(N+1,N+L)} = \begin{bmatrix} t(\mathbf{x}_{N+1}) \\ t(\mathbf{x}_{N+2}) \\ \vdots \\ t(\mathbf{x}_{N+L}) \end{bmatrix}, \quad \mathbf{t}_{(1,N+L)} = \begin{bmatrix} t(\mathbf{x}_1) \\ t(\mathbf{x}_2) \\ \vdots \\ t(\mathbf{x}_N) \\ t(\mathbf{x}_{N+1}) \\ t(\mathbf{x}_{N+2}) \\ \vdots \\ t(\mathbf{x}_{N+L}) \end{bmatrix}$$

Also, let $\mathbf{K}$ be a NxL sized matrix, where each element is $k(\mathbf{x}_n, \mathbf{x}_l)\,\forall\, n = 1, ..., N$ and $l = (N+1), ..., (N+L)$.

$\mathbf{C}_{(1,N)}$ be a NxN sized matrix, where each element is $k(\mathbf{x}_n, \mathbf{x}_m)\,\forall\, n = 1, ..., N$ and $m = 1, ..., N$.

$\mathbf{C}_{(N+1,N+L)}$ be a LxL sized matrix, where each element is $k(\mathbf{x}_l, \mathbf{x}_v)\,\forall\, l = (N+1), ..., (N+L)$ and $v = (N+1), ..., (N+L)$.

Then, the Gaussian prior is:

$$p(\mathbf{t}_{(1,N+L)}) = \mathcal{N}(\mathbf{0}, \mathbf{C}_{(1,N+L)})$$

$$\text{where} \quad \mathbf{C}_{(1,N+L)} = \begin{pmatrix} \mathbf{C}_{(1,N)} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{C}_{(N+1,N+L)} \end{pmatrix}$$

$$p(\mathbf{t}_{(1,N)}) = \mathcal{N}(\mathbf{0}, \mathbf{C}_{(1,N)})$$

$$p(\mathbf{t}_{(N+1,N+L)}) = \mathcal{N}(\mathbf{0}, \mathbf{C}_{(N+1,N+L)})$$

Using results 2.81 and 2.82, we get:

$$p\left(\mathbf{t}_{(N+1,N+L)}|\mathbf{t}_{(1,N)}\right) = \mathcal{N}\left(\boldsymbol{\mu}_{(N+1,N+L)|(1,N)}, \boldsymbol{\Sigma}_{(N+1,N+L)|(1,N)}\right)$$

where

$$\boldsymbol{\mu}_{(N+1,N+L)|(1,N)} = \mathbf{0} + \mathbf{K}^T\mathbf{C}_{(1,N)}^{-1}\mathbf{t}_{(1,N)}$$

$$= \mathbf{K}^T\mathbf{C}_{(1,N)}^{-1}\mathbf{t}_{(1,N)}$$

$$\boldsymbol{\Sigma}_{(N+1,N+L)|(1,N)} = \mathbf{C}_{(N+1,N+L)} - \mathbf{K}^T\mathbf{C}_{(1,N)}^{-1}\mathbf{K}$$

Based on these results, we can find the marginal distribution of the test observation t_j.

$\mathbf{K}^T$ is a LxN sized matrix, with every j^{th} row being a row vector has elements $k(x_r, x_n)$ for n = 1, . . . , N. So for one test observation t_j, we only need to consider $\mathbf{k}_r^T$ row vector.

Similarly, $\mathbf{C}_{(N+1,N+L)}$ is an LxL sized matrix, and we only need to consider the r x r^{th} element.

We finally get the results:

$$p(t_r|\mathbf{t}_{(1,N)}) = \mathcal{N}(\mu_{(r|(1,N))}, \sigma_{(r|(1,N))})$$

where

$$\mu_{(r|(1,N))} = \mathbf{k}_r^T\mathbf{C}_{(1,N)}^{-1}\mathbf{t}_{(1,N)}$$

and

$$\sigma_{(r|(1,N))}) = c_{r,r} - \mathbf{k}_r^T\mathbf{C}_{(1,N)}^{-1}\mathbf{k}_r$$

which are the usual Gaussian process regression results (6.66) and (6.67).