CHAPTER 5
Kernel Methods and Radial-Basis Function Networks

Problem 5.9

The expected square error is given by

$$J(F) = \frac{1}{2}\sum_{i=1}^{N} \int_{R^{m_0}} (f(\mathbf{x}_i) - F(\mathbf{x}_i, \xi))^2 f_{\xi}(\xi) d\xi$$

where $f_{\xi}(\xi)$ is the probability density function of a noise distribution in the input space R^{m_0}. It is reasonable to assume that the noise vector ξ is additive to the input data vector $\mathbf{x}$. Hence, we may define the cost function $J(F)$ as

$$J(F) = \frac{1}{2}\sum_{i=1}^{N} \int_{R^{m_0}} (f(\mathbf{x}_i) - F(\mathbf{x}_i + \xi))^2 f_{\xi}(\xi) d\xi \tag{1}$$

where (for convenience of presentation) we have interchanged the order of summation and integration, which is permissible because both operations are linear. Let

$$\mathbf{z} = \mathbf{x}_i + \xi \quad \text{or} \quad \xi = \mathbf{z} - \mathbf{x}_i$$

Hence, we may rewrite (1) in the equivalent form:

$$J(F) = \frac{1}{2} \int_{R^{m_0}} \sum_{i=1}^{N} (f(\mathbf{x}_i) - F(\mathbf{z}))^2 f_{\xi}(\mathbf{z} - \mathbf{x}_i) d\mathbf{z} \tag{2}$$

Note that the subscript ξ in $f_{\xi}(\cdot)$ merely refers to the "name" of the noise distribution and is therefore untouched by the change of variables. Differentiating (2) with respect to F, setting the result equal to zero, and finally solving for $F(\mathbf{z})$, we get the optimal estimator

$$\hat{F}(\mathbf{z}) = \frac{\sum_{i=1}^{N} f(\mathbf{x}_i) f_{\xi}(\mathbf{z} - \mathbf{x}_i)}{\sum_{i=1}^{N} f_{\xi}(\mathbf{z} - \mathbf{x}_i)}$$

This result bears a close resemblance to the Watson-Nadaraya estimator.