3.3

If we define $R = diag(r_1, ..., r_N)$ to be a diagonal matrix containing the weighting coefficients, then we can write the weighted sum-of-squares cost function in the form

$E_D(w) = \frac{1}{2}(t - \Phi w)^T R(t - \Phi w)$

Setting the derivative with respect to w to zero, and re-arranging, then gives

$W^* = (\Phi^T R \Phi)^{-1} \Phi^T Rt$

which reduces to the standard solution for the case R = I

from the book $p(t|X,\ w,\ \beta)\ =\ \prod_{n=1}^{N} N\ (t_n|w^T \varphi(x_n),\ \beta^{-1})$

$lnp(t|X,\ w,\ \beta) = \frac{N}{2} ln\beta - \frac{N}{2} ln(2\pi) - \beta E_D(w)$

$E_D(w)\ = \frac{1}{2}\ \sum_{n=1}^{N}\ \{t_n - W^T \varphi(x_n)\}^2$ If we compare with these functions, we see that r_n can be regarded as a precision (inverse variance) parameter, particular to the data point(x_n, t_n), that either replaces or scales

Alternatively, r_n can be regarded as an effective number of replicated observations of data point (x_n, t_n); this becomes particularly clear if we consider the function problem with r_n taking positive integer values, although it is valid for any $r_n > 0$.

3.11

from the book function

$\sigma_N^2 = \frac{1}{\beta} + \Phi(X)^T S_N \Phi(X)$

we can have

$\sigma_{N+1}^2 = \frac{1}{\beta} + \Phi(X)^T S_{N+1} \Phi(X)$

and can get

$S_{N+1}\ =\ (S_N^{-1}\ +\ \beta \Phi_{N+1} {\Phi^T}_{N+1}\ )^{-1}$

$= S_N - \frac{(S_N \Phi_{N+1}\ \beta^{\frac{1}{2}})(\beta^{\frac{1}{2}} {\Phi^T}_{N+1} S_N)}{1 + \beta {\Phi^T}_{N+1} S_N \Phi_{N+1}}$

$= S_N - \frac{(\beta S_N \Phi_{N+1}\ {\Phi^T}_{N+1} S_N)}{1 + \beta {\Phi^T}_{N+1} S_N \Phi_{N+1}}$

then

${\sigma^2}_{N+1}(X) = \frac{1}{\beta} + \Phi(X)^T (S_N - \frac{\beta S_N \Phi_{N+1}\ {\Phi^T}_{N+1} S_N}{1 + \beta {\Phi^T}_{N+1} S_N \Phi_{N+1}}) \Phi(X)$

$= {\sigma^2}_N(X) - \frac{\beta \Phi(X)^T S_N \Phi_{N+1}\ {\Phi^T}_{N+1} S_N \Phi(X)}{1 + \beta {\Phi^T}_{N+1} S_N \Phi_{N+1}}$

Since SN is positive definite, the numerator and denominator of the second term will be non-negative and positive, respectively, and hence $\sigma_{N+1}^2(x) \leq \sigma_N^2(x)$

3.14

For $\alpha = 0$ the covariance matrix S_N becomes:

$S_N = (\beta \Phi^T \Phi)^{-1}$

Let us define a new set of orthonormal basis functions given by linear combinations of the original basis functions so that

$\psi(x) = V \phi(x)$

where V is an M M matrix. Since both the original and the new basis functions

are linearly independent and span the same space, this matrix must be invertible and hence
$\phi(x) = V^{-1}\psi(x)$
For the data set x_n
$\Psi = \Phi V^T$
and consequently $\Psi = \Phi V^{-T}$
where V_T denotes $(V^1)^T$. Orthonormality implies $\Psi^T\Psi = I$
the covariance matrix then becomes
$S_N = \beta^{-1}(\Phi^T\Phi)^{-1} = \beta^{-1}(V^{-T}\Psi^T\Psi V^{-1})^{-1} = \beta^{-1}V^TV$
Here we have used the orthonormality of the i(x). Hence the equivalent kernel becomes
$k(x, x') = \beta\phi(x)^T S_N \phi(x') = \phi(x)^T V^T V\phi(x') = \psi(x)^T\psi(x')$
as required. From the orthonormality condition, and setting j = 1, it follows that
$\sum_{n=1}^{N} \psi_i(x_n)\ \psi_1(x_n)\ = \sum_{n=1}^{N} \psi_i(x_n) =\ \delta_{i1}$
where we have used $\psi_1(x) = 1$. Now consider the sum
$\sum_{n=1}^{N} k(x, x_n) = \sum_{n=1}^{N} \psi(x_n)\ ^T\psi(x_n)$
$= \sum_{n=1}^{N} \sum_{i=1}^{M} \psi_i(x)\psi_i(x_n)$
$=\ \sum_{i=1}^{M} \psi_i(x)\delta_{i1} = \psi_1(x) = 1$
which proves the summation constraint as required.

3.21
where u_i are a set of M orthonormal vectors, and the M eigenvalues $_i$ are all real. We first express the left hand side in terms of the eigenvalues of A. The log of the determinant of A can be written as
$\ln|A| = \ln\prod_{i=1}^{M}\eta_i = \sum_{i=1}^{M}\ln\ \eta_i$
Taking the derivative with respect to some scalar α we obtain
$\frac{d}{d\alpha}\ln|A| = \sum_{i=1}^{M}\ \frac{1}{\eta_i}\frac{d}{d\alpha}\ \eta_i$

We now express the right hand side in terms of the eigenvector expansion and show that it takes the same form. First we note that A can be expanded in terms of its own eigenvectors to give
$A = \sum_{i=1}^{M}\ \eta_i u_i u_i^T$
and similarly the inverse can be written as
$A^{-1} = \sum_{i=1}^{M}\ \frac{1}{\eta_i} u_i u_i^T$
then
$Tr(A^{-1}\frac{d}{d\alpha}A) = Tr(\sum_{i=1}^{M}\ \frac{1}{\eta_i} u_i u_i^T \frac{d}{d\alpha}\sum_{j=1}^{M}\ \eta_j u_j u_j^T)$
$= Tr(\sum_{i=1}^{M}\ \frac{1}{\eta_i} u_i u_i^T \{\sum_{j=1}^{M} \frac{d\eta_j}{d\alpha} u_j u_j^T + \eta_j(b_j u_j^T + u_j b_j^T)\})$
$= Tr(\sum_{i=1}^{M}\ \frac{1}{\eta_i} u_i u_i^T \sum_{j=1}^{M} \frac{d\eta_j}{d\alpha} u_j u_j^T)\ + Tr(\sum_{i=1}^{M}\ \frac{1}{\eta_i} u_i u_i^T \sum_{j=1}^{M} \eta_j(b_j u_j^T + u_j b_j^T))$
where $b_j = \frac{du_j}{/}d\alpha$. Using the properties of the trace and the orthognality of eigenvectors, we can rewrite the second term as

$Tr(\sum_{i=1}^{M} \frac{1}{\eta_i} u_i u_i^T \sum_{j=1}^{M} \eta_j (b_j u_j^T + u_j b_j^T)) = Tr(\sum_{i=1}^{M} \frac{1}{\eta_i} u_i u_i^T \sum_{j=1}^{M} 2\eta_j u_j b_j^T)$
$= Tr(\sum_{i=1}^{M} \sum_{j=1}^{M} \frac{2\eta_j}{\eta_i} u_i u_i^T u_j b_j^T)$
$= Tr(\sum_{i=1}^{M} b_j u_j^T + u_j b_j^T)$
$= Tr(\frac{d}{d\alpha} \sum_{i=1}^{M} u_i u_i^T)$
and
$\sum_{i=1}^{M} u_i u_i^T = I$
which is constant and thus its derivative w.r.t. will be zero
We again use the properties of the trace and the orthognality of eigenvectors to obtain
$Tr(A^{-1} \frac{d}{d\alpha} A) = \sum_{i=1}^{M} \frac{1}{\eta_i} \frac{d\eta_i}{d\alpha}$
We have now shown that both the left and right hand sides take the same form when expressed in terms of the eigenvector expansion. Next, we use (3.117) to differentiate (3.86) w.r.t. , yielding
$\frac{d}{d\alpha} \ln\ p(t|\alpha\ \beta) = \frac{M}{2}\frac{1}{\alpha} - \frac{1}{2} m_N^T m_N - \frac{1}{2} Tr(A^{-1} \frac{d}{d\alpha} A)$
$= \frac{1}{2}(\frac{M}{\alpha} - m_N^T m_N - Tr(A^{-1}))$
$= \frac{1}{2}(\frac{M}{\alpha} - m_N^T m_N - \sum_i \frac{1}{\lambda_i - \alpha})$
which we recognize as the r.h.s. of (3.89), from which (3.92) can be derived as de- tailed in Section 3.5.2, immediately following (3.89).