3.5

From constraint (3.30) we have $\sum_{j=1}^{M} |w_j|^2 \leq \eta$. And this can be re-written to $\sum_{j=1}^{M} |w_j|^2 - \eta \leq 0 \Rightarrow \frac{1}{2}(\sum_{j=1}^{M} |w_j|^2 - \eta \leq 0)$.

With the help of Appendix E, the Lagrange function is defined by $L(x, \lambda) \equiv f(x) + \lambda g(x)$, with $g(x)$ being the constraint equation.

Hence, the Lagrange function in this case can be written as

$$L(w, \lambda) = E_D(w) + \lambda g(x)$$
$$= \frac{1}{2} \sum_{n=1}^{N} \{t_n - w^T \phi(x_n)\}^2 + \frac{\lambda}{2} (\sum_{j=1}^{M} |w_j|^2 - \eta)$$

We can see that this Lagrange function is very much the same with (3.29), and they have the same dependence on w. Hence, the minimization of function (3.29) is equivalent to minimizing (3.12) subject to the constraint (3.30).

For the relationship between the parameters η and λ, suppose we have minimized the Lagrange function $L(w, \lambda)$, and denote the resulting w with $\overline{w}(\lambda)$, with the help of equation (E.11) in Appendix E, we would have $\lambda g(x) = 0$

$$\lambda \cdot \frac{1}{2} (\sum_{j=1}^{M} |\overline{w}(\lambda)|^2 - \eta) = 0$$
$$\frac{\lambda}{2} \sum_{j=1}^{M} |\overline{w}(\lambda)|^2 = \frac{\lambda}{2} \eta$$
$$\eta = \sum_{j=1}^{M} |\overline{w}(\lambda)|^2$$

3.11

With the help of (3.51) and Exercise 3.8, we have

$$S_{N+1}^{-1} = S_N^{-1} + \beta\phi(x_{N+1})\phi(x_{N+1})^T = S_N^{-1} + \sqrt{\beta}\,\phi(x_{N+1})\sqrt{\beta}\,\phi(x_{N+1})^T$$

With the help of equation (3.110), which is simply a special case of the Woodbury identity in Appendix C, if we identify S_N^{-1} with M and $\sqrt{\beta}\,\phi(x_{N+1})$ with V, we would have

$$(S_{N+1}^{-1})^{-1} = S_{N+1} = S_N - \frac{(S_N(\sqrt{\beta}\phi(x_{N+1})))((\sqrt{\beta}\phi(x_{N+1}))^T S_N)}{1+(\sqrt{\beta}\phi(x_{N+1}))^T S_N(\sqrt{\beta}\phi(x_{N+1}))} = S_N - \frac{\beta(S_N\phi(x_{N+1}))((\phi(x_{N+1}))^T S_N}{1+\beta(\phi(x_{N+1}))^T S_N(\phi(x_{N+1}))}$$

Since the claim of (3.111) is $\sigma_{N+1}^2(x) \leq \sigma_N^2(x)$, and with the use of (3.59), we have

$$\begin{aligned}
\sigma_N^2(x) - \sigma_{N+1}^2(x) &= \frac{1}{\beta} + \phi(x)^T S_N \phi(x) - \left(\frac{1}{\beta} + \phi(x)^T S_{N+1}\phi(x)\right) \\
&= \phi(x)^T S_N \phi(x) - \phi(x)^T S_{N+1}\phi(x) \\
&= \phi(x)^T(S_N - S_{N+1})\phi(x) \\
&= \phi(x)^T\left(S_N - S_N - \frac{\beta(S_N\phi(x_{N+1}))((\phi(x_{N+1}))^T S_N)}{1+\beta(\phi(x_{N+1}))^T S_N(\phi(x_{N+1}))}\right)\phi(x) \\
&= \phi(x)^T \frac{\beta(S_N\phi(x_{N+1}))((\phi(x_{N+1}))^T S_N)}{1+\beta(\phi(x_{N+1}))^T S_N(\phi(x_{N+1}))}\phi(x) \\
&= \phi(x)^T \frac{(S_N\phi(x_{N+1}))((\phi(x_{N+1}))^T S_N)}{1/\beta+(\phi(x_{N+1}))^T S_N(\phi(x_{N+1}))}\phi(x) \\
&= \frac{\phi(x)^T(S_N\phi(x_{N+1}))((\phi(x_{N+1}))^T S_N)\phi(x)}{1/\beta+(\phi(x_{N+1}))^T S_N(\phi(x_{N+1}))} \\
&= \frac{[\phi(x)^T S_N\phi(x_{N+1})]^2}{1/\beta+(\phi(x_{N+1}))^T S_N(\phi(x_{N+1}))}
\end{aligned}$$

With S_N being positive semi-definite, both the numerator and denominator of the final result will be positive. So, we have $\sigma_N^2(x) - \sigma_{N+1}^2(x) \geq 0$ and hence $\sigma_{N+1}^2(x) \leq \sigma_N^2(x)$.

4.9

So we have $P(C_k) = \pi_k$ and $P(\phi_n, C_k) = P(\phi_n|C_k)P(C_k)$. The probability of a single data point is $P(\phi_n, t_n|\pi) = \prod_{k=1}^{K} [\pi_k P(\phi_n|C_k)]^{t_{nk}}$

And the likelihood function is $P(\phi_n, t_n|\{\pi_k\}) = \prod_{n=1}^{N}\prod_{k=1}^{K} [\pi_k P(\phi_n|C_k)]^{t_{nk}}$

Now assuming we have i.i.d. data, the logarithm likelihood becomes

$$\ln P(\phi_n, t_n|\{\pi_k\}) = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}[\ln \pi_k + \ln P(\phi_n|C_k)]$$

So we want to maximize this logarithm likelihood function with respect to π, by adding a Lagrange Multiplier to it, we have

$$L = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \ln \pi_k + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right)$$

By differentiating this function with respect to π_k, we have

$$\frac{\partial L}{\partial \pi_k} = \sum_{n=1}^{N} \frac{t_{nk}}{\pi_k} + \lambda = \frac{N_k}{\pi_k} + \lambda = 0$$

$$\frac{N_k}{\pi_k} = -\lambda$$

$$\pi_k = -\frac{N_k}{\lambda}$$

And by summing both sides over k, we have

$$\sum_{k=1}^{K} \pi_k = -\frac{1}{\lambda}\sum_{k=1}^{K} N_k = 1$$

$$\lambda = -N$$

So, $\pi_k = -\frac{N_k}{\lambda} = -\frac{N_k}{-N} = \frac{N_k}{N}$, and (4.159) is proved.

4.20

With the equation given by 4.110, we have

$$\nabla_{w_k}\nabla_{w_j}E(w_1,\cdots,w_k) = -\sum_{n=1}^{N} y_{nk}(I_{kj}-y_{nj})\phi_n\phi_n^T,$$

and this equation denotes the $(k,j)^{th}$ block of size $M\times M$. And we can obtain that

$$u^T H u = \sum_{k=1}^{K}\sum_{j=1}^{K} u_k^T H_{kj} u_j,$$

and H_{kj} is the $(k,j)^{th}$ block matrix of the $M\times M$ matrix H.

With the help of (4.110), we have

$$\sum_{k=1}^{K}\sum_{j=1}^{K} u_k^T H_{kj} u_j = \sum_{k=1}^{K}\sum_{j=1}^{K} u_k^T\left[\sum_{n=1}^{N} y_{nk}(I_{kj}-y_{nj})\phi_n\phi_n^T\right]u_j$$

$$= \sum_{n=1}^{N}\left[\sum_{k=1}^{K} y_{nk}u_k^T\phi_n\phi_n^T\sum_{j=1}^{K}(I_{kj}-y_{nj})u_j\right]$$

$$= \sum_{n=1}^{N}\left[\sum_{k=1}^{K} y_{nk}u_k^T\phi_n\phi_n^T\left[\sum_{j=1}^{K}(I_{kj}-y_{nj})u_j\right]\right]$$

$$= \sum_{n=1}^{N}\left[\sum_{k=1}^{K} y_{nk}u_k^T\phi_n\phi_n^T\left[\sum_{j=1}^{K}I_{kj}u_j - \sum_{j=1}^{K}y_{nj}u_j\right]\right]$$

$$= \sum_{n=1}^{N}\left[\sum_{k=1}^{K} y_{nk}u_k^T\phi_n\phi_n^T\left[u_k - \sum_{j=1}^{K}y_{nj}u_j\right]\right]$$

$$= \sum_{n=1}^{N}\left[\sum_{k=1}^{K} y_{nk}u_k^T\phi_n\phi_n^T u_k - \sum_{k=1}^{K} y_{nk}u_k^T\phi_n\phi_n^T\sum_{j=1}^{K}y_{nj}u_j\right]$$

Now define a positive function

$$\begin{aligned} f(u) &= u^T\phi_n\phi_n^T u \\ &= (u^T\phi_n)(\phi_n^T u) \\ &= (u^T\phi_n)(u^T\phi_n) \\ &= (u^T\phi_n)^2 \geqslant 0 \end{aligned}$$

So it's positive semi-definite.

Now,

$$\sum_{n=1}^{N}\left[\sum_{k=1}^{K} y_{nk} u_k^T \phi_n \phi_n^T u_k - \sum_{k=1}^{K} y_{nk} u_k^T \phi_n \phi_n^T \sum_{j=1}^{K} y_{nj} u_j\right]$$

$$= \sum_{n=1}^{N}\left[\sum_{k=1}^{K} y_{nk} f(u_k) - f\left(\sum_{k=1}^{K} y_{nk} u_k\right)\right]$$

With the function $f(u)$ being positive semi-definite, and $\sum_{k=1}^{K} y_{nk} = 1$, and the use of Jensen's inequality: $f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \le \sum_{i=1}^{M} \lambda_i f(x_i)$, we can have

$$f\left(\sum_{k=1}^{K} y_{nk} u_k\right) \le \sum_{k=1}^{K} y_{nk} f(u_k)$$

And this means that each and every term in the sum over n is non-negative, and hence, $\sum_{k=1}^{K}\sum_{j=1}^{K} u_k^T H_{kj} u_j \ge 0$, and (4.110) is positive semi-definite.