1 5.9

$$
\begin{aligned}
\mathbf{S} &= \mathbf{N}(\mathbf{N}^T\mathbf{N} + \lambda\Omega_N)^{-1}\mathbf{N}^T \\
&= \mathbf{N}(\mathbf{N}^T(\mathbf{I} + \lambda(\mathbf{N}^T)^{-1}\Omega_N\mathbf{N}^{-1})\mathbf{N})^{-1}\mathbf{N}^T \\
&= (\mathbf{I} + \lambda\mathbf{K})^{-1}
\end{aligned}
$$

其中$\mathbf{K} = (\mathbf{N}^T)^{-1}\Omega_N N^{-1}$

2 5.13

记$\hat{f}_\lambda^{(-i)}$是排除第i个样本之后的拟合函数。

$$
\begin{aligned}
\hat{f}_\lambda^{(-i)}(x_i) &= \frac{1}{1 - S_\lambda(i,i)} \sum_{j\neq i} S_\lambda(i,j)y_j \\
&= \sum_{j\neq i} S_\lambda(i,j)y_j + S_\lambda(i,i)\hat{f}_\lambda^{(-i)}(x_i)
\end{aligned}
$$

也就是说

$$
\begin{aligned}
\hat{f}_\lambda^{(-i)}(x_i) &= \hat{f}_\lambda(x_i) + S_\lambda(i,i)\hat{f}_\lambda^{(-i)}(x_i) - S_\lambda(i,i)y_i \\
(1 - S_\lambda(i,i))\left(\hat{f}_\lambda^{(-i)}(x_i) - y_i\right) &= \hat{f}_\lambda(x_i) - y_i \\
y_i - \hat{f}_\lambda^{(-i)}(x_i) &= \frac{y_i - \hat{f}_\lambda(x_i)}{1 - S_\lambda(i,i)}
\end{aligned}
$$

所以

$$
CV(\hat{f}_\lambda) = \sum_{i=1}^{N}\left(y_i - \hat{f}_\lambda^{(-i)}(x_i)\right)^2 = \sum_{i=1}^{N}\left(\frac{y_i - \hat{f}_\lambda(x_i)}{1 - S_\lambda(i,i)}\right)^2
$$

3 5.15

(1)

$$
\begin{aligned}
\langle K(\cdot, x_i), f\rangle_{\mathcal{H}_K} &= \left\langle \sum_{j=1}^{\infty}\gamma_j\phi_j(x)\phi_j(x_i), \sum_{j=1}^{\infty}c_j\phi_j(x)\right\rangle \\
&= \sum_{j=1}^{\infty}\frac{c_j\gamma_j\phi_j(x_i)}{\gamma_j} \\
&= f(x_i)
\end{aligned}
$$

(2)

$$
\begin{aligned}
\langle K(\cdot, x_i), K(\cdot, x_j)\rangle_{\mathcal{H}_K} &= \left\langle \sum_{k=1}^{\infty}\gamma_k\phi_k(x)\phi_k(x_i), \sum_{k=1}^{\infty}\gamma_k\phi_k(x)\phi_k(x_j)\right\rangle \\
&= \sum_{k=1}^{\infty}\frac{\gamma_k^2\phi_k(x_i)\phi_k(x_j)}{\gamma_k} \\
&= K(x_i, x_j)
\end{aligned}
$$

(3)

$$
\begin{aligned}
J(g) &= \left\langle \sum_{i=1}^{N}\alpha_i K(x, x_i), \sum_{i=1}^{N}\alpha_i K(x, x_i)\right\rangle \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}K(x_i, x_j)\alpha_i\alpha_j
\end{aligned}
$$

(4)

$$
\begin{aligned}
J(g) &= \left\langle \sum_{i=1}^{N} \alpha_i K(x, x_i), \sum_{i=1}^{N} \alpha_i K(x, x_i) \right\rangle \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} K(x_i, x_j) \alpha_i \alpha_j
\end{aligned}
$$

(5)

$$
\begin{aligned}
J(\tilde{g}) &= J(g) + 2\langle J(g), \rho \rangle + \|\rho\|_{\mathcal{H}_K}^2 \\
&= J(g) + \|\rho\|_{\mathcal{H}_K}^2 \\
&\geq J(g)
\end{aligned}
$$

由第(1)问的结论可知

$$
\begin{aligned}
\tilde{g}(x_i) &= \langle K(\cdot, x_i), \tilde{g} \rangle_{\mathcal{H}_K} \\
&= \langle K(\cdot, x_i), g + \rho \rangle_{\mathcal{H}_K} \\
&= \langle K(\cdot, x_i), g \rangle_{\mathcal{H}_K}
\end{aligned}
$$

所以

$$
L(y_i, \tilde{g}(x_i)) = L(y_i, g(x_i))
$$

于是

$$
\sum_{i=1}^{N} L(y_i, \tilde{g}(x_i)) + \lambda J(\tilde{g}) \geq \sum_{i=1}^{N} L(y_i, g(x_i)) + \lambda J(g)
$$

当且仅当$\rho(x) = 0$时，上式取等。

4 5.16

(1) 由核函数定义

$$
K(x, y) = \sum_{m=1}^{M} h_m(x) h_m(y) = \sum_{i=1}^{\infty} \gamma_i \phi_i(x) \phi_i(y)
$$

$$
\sum_{m=1}^{M} \langle h_m(x), \phi_k(x) h_m(y) = \sum_{i=1}^{\infty} \langle \phi_i(x), \phi_k(x) \phi_i(y)
$$

由于$\phi_i(x)$之间互相正交

$$
\sum_{m=1}^{M} \langle h_m(x), \phi_k(x) h_m(y) = \gamma_k \phi_k(y)
$$

令$g_{km} = \int h_m(x) \phi_k(x) dx$，那么

$$
\sum_{m=1}^{M} g_{km} h_m(y) = \gamma_k \phi_k(y)
$$

$$
\sum_{m=1}^{M} g_{km} g_{lm} = \gamma_k \delta_{k,l} = \sqrt{\gamma_k} \sqrt{\gamma_l} \delta_{kl}
$$

记$\mathbf{G} = \{g_{nm}\} \in \mathbb{R}^{M \times N}$，$\mathbf{V}^T = \mathbf{D}_\gamma^{-\frac{1}{2}} \mathbf{G}$，有

$$
\mathbf{G}\mathbf{G}^T = \text{diag}\{\gamma_1, \gamma_2, \ldots, \gamma_M\} = \mathbf{D}_\gamma
$$

$$
\mathbf{V}\mathbf{V}^T = I
$$

所以$\sum_{m=1}^{M} g_{km} h_m(y) = \gamma_k \phi_k(y)$可以写成$\mathbf{D}_\gamma \phi(y) = \mathbf{G} h(y)$

$$h(x)=\mathbf{V}\mathbf{V}^T h(x)=\mathbf{V}\mathbf{D}_\gamma^{-\frac{1}{2}}\mathbf{G}h(x)=\mathbf{V}\mathbf{D}_\gamma^{\frac{1}{2}}\phi(x)$$

令$\beta=(\beta_1,\beta_2,\ldots,\beta_m)^T$，$c=\mathbf{D}_\gamma^{\frac{1}{2}}\mathbf{V}^T\beta$

$$\begin{aligned}
&\min_{\{\beta_m\}_1^M}\sum_{i=1}^N\left(y_i-\sum_{m=1}^M\beta_m h_m(x_i)\right)^2+\lambda\sum_{m=1}^M\beta_m^2\\
=&\min_\beta\sum_{i=1}^N(y_i-\beta^T h(x_i))^2+\lambda\beta^T\beta\\
=&\min_\beta\sum_{i=1}^N(y_i-\beta^T\mathbf{V}\mathbf{D}_\gamma^{\frac{1}{2}}\phi(x_i))^2+\lambda\beta^T\beta\\
=&\min_c\sum_{i=1}^N(y_i-c^T\phi(x_i))^2+\lambda(\mathbf{V}\mathbf{D}_\gamma^{\frac{1}{2}}c)^T\mathbf{V}\mathbf{D}_\gamma^{\frac{1}{2}}c\\
=&\min_c\sum_{i=1}^N(y_i-c^T\phi(x_i))^2+\lambda c^T c\mathbf{D}_\gamma^{-1}\\
=&\min_{\{c_j\}_1^\infty}\sum_{i=1}^N\left(y_i-\sum_{j=1}^\infty c_j\phi_j(x_i)\right)^2+\lambda\sum_{j=1}^\infty\frac{c_j^2}{\gamma_j}
\end{aligned}$$

(2)

对于$\min_\beta\sum_{i=1}^N(y_i-\beta^T h(x_i))^2+\lambda\beta^T\beta$，解为$\hat{\beta}=(\mathbf{H}^T\mathbf{H}+\lambda\mathbf{I})^{-1}\mathbf{H}^T$。

于是

$$\begin{aligned}
\hat{\mathbf{f}}&=\mathbf{H}(\mathbf{H}^T\mathbf{H}+\lambda\mathbf{I})^{-1}\mathbf{H}^T\\
&=\frac{1}{\lambda}\mathbf{H}\mathbf{H}^T-\frac{1}{\lambda}\mathbf{H}\mathbf{H}^T\left(\lambda\mathbf{I}+\mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{H}\mathbf{H}^T\\
&=\frac{1}{\lambda}\mathbf{H}\mathbf{H}^T\left[(\lambda\mathbf{I}+\mathbf{H}\mathbf{H}^T)^{-1}(\lambda\mathbf{I}+\mathbf{H}\mathbf{H}^T)-(\lambda\mathbf{I}+\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{H}^T\right]\\
&=\frac{1}{\lambda}\mathbf{H}\mathbf{H}^T\left[(\lambda\mathbf{I}+\mathbf{H}\mathbf{H}^T)^{-1}\lambda\mathbf{I}\right]\\
&=\mathbf{H}\mathbf{H}^T(\lambda\mathbf{I}+\mathbf{H}\mathbf{H}^T)^{-1}
\end{aligned}$$

所以$\hat{\mathbf{f}}=\mathbf{K}(\mathbf{K}+\lambda\mathbf{I})^{-1}$。

（3）由上一问的结论，知

$$\begin{aligned}
\hat{f}(x)&=h(x)^T\hat{\beta}\\
&=\sum_{i=1}^N K(x,x_i)(\mathbf{K}+\lambda\mathbf{I})^{-1}\mathbf{y}
\end{aligned}$$

带入$\hat{\alpha}=(\mathbf{K}+\lambda\mathbf{I})^{-1}\mathbf{y}$即可。

（4）$\lambda\neq 0$时上述结论依然成立。若$\lambda=0$，$\hat{\mathbf{f}}=\mathbf{y}$。