Homework of Chapter 5

Chen Cheng, 1130339005

January 12, 2015

Ex. 5.9

Solution.
Let $\mathbf{N} = \mathbf{UDV}^T$, and we have

$$\begin{aligned}\mathbf{S}_\lambda =&\mathbf{UDV}^T(\mathbf{VDU}^T\mathbf{UDV}^T + \lambda\mathbf{\Omega}_N)^{-1}(\mathbf{UDV}^T)^T\\ =&\mathbf{U}(\mathbf{D}^{-1}\mathbf{V}^T\mathbf{VD}^2\mathbf{V}^T\mathbf{VD}^{-1} + \lambda\mathbf{D}^{-1}\mathbf{V}^T\mathbf{\Omega}_N\mathbf{VD}^{-1})^{-1}\mathbf{U}^T\\ =&(\mathbf{UU}^T + \lambda\mathbf{UD}^{-1}\mathbf{V}^T\mathbf{\Omega}_N\mathbf{VD}^{-1}\mathbf{U}^T)^{-1}\\ =&(\mathbf{I} + \lambda\mathbf{UD}^{-1}\mathbf{V}^T\mathbf{\Omega}_N\mathbf{VD}^{-1}\mathbf{U}^T)^{-1}\\ =&(\mathbf{I} + \lambda\mathbf{K})^{-1}\end{aligned}$$

This the the Reinsch form for the smoothing spline. □

Ex. 5.13

Solution. We assume that $\hat{f}'_\lambda(x)$ is the new smoothing spline trained by $N+1$ samples: $(x_0, \hat{f}_\lambda^{(-0)}(x_0)), (x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$, and $\hat{f}_\lambda^{(-0)}(x)$ represents the old smoothing spline trained by N pairs $(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$. We can find that $\hat{f}_\lambda^{(-0)}$ is the solution of the following problem:

$$\min_f RSS(f) = \min_f \sum_{i=1}^{N}\{y_i - f(x_i)\}^2 + \lambda\int\{f''(t)\}^2 \mathrm{d}t$$

and $\hat{f}'_\lambda$ is the solution of the following problem:

$$\min_f RSS(f) = \min_f \sum_{i=1}^{N}\{y_i - f(x_i)\}^2 + (\hat{f}_\lambda^{(-0)}(x_0) - f(x_0))^2 + \lambda\int\{f''(t)\}^2 \mathrm{d}t$$

Then

$$\begin{aligned}
&\sum_{i=1}^{N}\{y_i - f(x_i)\}^2 + (\hat{f}_\lambda^{(-0)}(x_0) - f(x_0))^2 + \lambda \int \{f''(t)\}^2 \mathrm{d}t \\
\geq &\sum_{i=1}^{N}\{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 \mathrm{d}t \\
\geq &\sum_{i=1}^{N}\{y_i - \hat{f}_\lambda^{(-0)}(x_i)\}^2 + \lambda \int \{\hat{f}_\lambda''^{(-0)}(t)\}^2 \mathrm{d}t \\
= &\sum_{i=1}^{N}\{y_i - \hat{f}_\lambda^{(-0)}(x_i)\}^2 + (\hat{f}_\lambda^{(-0)}(x_0) - \hat{f}_\lambda^{(-0)}(x_0))^2 + \lambda \int \{\hat{f}_\lambda''^{(-0)}(t)\}^2 \mathrm{d}t
\end{aligned}$$

with equality iff $f(x_i) = \hat{f}_\lambda^{(-0)}(x_i)$ for $i = 0, 1, ..., N$
So, $\hat{f}'_\lambda(x_i) = \hat{f}_\lambda^{(-0)}(x_i)$ for $i = 0, 1, ..., N$
Now, we begin to derive the formula (5.26)
Let

$$y'_i = \begin{cases} \hat{f}_\lambda^{(-0)}(x_0) & i = 0 \\ y_i & i \neq 0 \end{cases}$$

As $\mathbf{f'} = \mathbf{S}'_\lambda \mathbf{y'}$, We have

$$\hat{f}_\lambda^{(-0)}(x_0) = \hat{f}'_\lambda(x_0) = \sum_{i=0}^{N} \mathbf{S}'_\lambda(0, i) y'_i$$

Let $\hat{f}_\lambda(x) = \mathbf{S}_\lambda \mathbf{y}$ is the smoothing spline trained by $N + 1$ samples: $(x_0, y_0), (x_1, y_1), ..., (x_N, y_N)$.
As $\mathbf{S}_\lambda$ is only depend on x_i and λ, we have $\mathbf{S}'_\lambda = \mathbf{S}_\lambda$.

Then,

$$\begin{aligned}
y_0 - \hat{f}_\lambda(x_0) =& y_0 - \sum_{i=0}^{N} \mathbf{S}_\lambda(0,i) y_i \\
=& y_0 - \mathbf{S}_\lambda(0,0) y_0 - \sum_{i=1}^{N} \mathbf{S}_\lambda(0,i) y_i \\
=& y_0 - \mathbf{S}_\lambda(0,0) y_0 - \sum_{i=1}^{N} \mathbf{S}'_\lambda(0,i) y'_i \\
=& y_0 - \mathbf{S}_\lambda(0,0) y_0 - \left[\sum_{i=0}^{N} \mathbf{S}'_\lambda(0,i) y'_i - \mathbf{S}'_\lambda(0,0) y'_0 \right] \\
=& y_0 - \mathbf{S}_\lambda(0,0) y_0 - \left[\hat{f}_\lambda^{(-0)}(x_0) - \mathbf{S}_\lambda(0,0) \hat{f}_\lambda^{(-0)}(x_0) \right] \\
=& [1 - \mathbf{S}_\lambda(0,0)][y_0 - \hat{f}_\lambda^{(-0)}(x_0)]
\end{aligned}$$

So,

$$y_0 - \hat{f}_\lambda^{(-0)}(x_0) = \frac{y_0 - \hat{f}_\lambda(x_0)}{1 - \mathbf{S}_\lambda(0,0)}$$

Similarly, we can get

$$y_i - \hat{f}_\lambda^{(-i)}(x_i) = \frac{y_i - \hat{f}_\lambda(x_i)}{1 - \mathbf{S}_\lambda(i,i)} \quad i = 0, 1, ..., N$$

From this, we have

$$\begin{aligned}
CV(\hat{f}_\lambda) =& \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{f}_\lambda^{(-i)})^2 \\
=& \frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_i - \hat{f}_\lambda(x_i)}{1 - \mathbf{S}_\lambda(i,i)} \right)^2
\end{aligned}$$

$\square$

Ex. 5.15

Solution. Suppose $f(x) = \sum_{i=1}^{\infty} c_i\phi_i(x)$, $h(x) = \sum_{i=1}^{\infty} d_i\phi_i(x)$, and $K(x, y) = \sum_{i=1}^{\infty} \gamma_i\phi_i(x)\phi_i(y)$.

According to the property of RKHS, we have

$$
\begin{aligned}
\langle f(x), h(x)\rangle_{\mathcal{H}_K} =& \frac{1}{4}(\|f(x)+h(x)\|^2_{\mathcal{H}_K} - \|f(x)-h(x)\|^2_{\mathcal{H}_K}) \\
=& \frac{1}{4}\sum_{i=1}^{\infty} \frac{(c_i+d_i)^2 + (c_i-d_i)^2}{\gamma_i} \\
=& \sum_{i=1}^{\infty} \frac{c_i d_i}{\gamma_i}
\end{aligned}
$$

(a) Using the property of RKHS, we have

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

(b) Using the property of RKHS, we have

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

(c) Using the result of (b), we have

$$
\begin{aligned}
J(g(x)) =&\langle g(x), g(x)\rangle_{\mathcal{H}_K} \\
=&\langle \sum_{i=1}^{N} \alpha_i K(x, x_i), \sum_{j=1}^{N} \alpha_j K(x, x_j)\rangle_{\mathcal{H}_K} \\
=&\sum_{i=1}^{N}\sum_{j=1}^{N}\langle K(x, x_i), K(x, x_j)\rangle_{\mathcal{H}_K}\alpha_i\alpha_j \\
=&\sum_{i=1}^{N}\sum_{j=1}^{N} K(x_i, x_j)\alpha_i\alpha_j
\end{aligned}
$$

(d) Using the result of (a) and the fact that ρ is orthogonal to each of $K(x, x_i)$, we have

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

Moreover, we have

$$
\begin{aligned}
\langle g, \rho\rangle =&\langle \sum_{i=1}^{N} \alpha_i K(x, x_i), \rho\rangle_{\mathcal{H}_K} \\
=&\sum_{i=1}^{N} \alpha_i \langle K(x, x_i), \rho\rangle_{\mathcal{H}_K} \\
=&0
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
J(\widetilde{g}) =&\langle g + \rho, g + \rho\rangle_{\mathcal{H}_K} \\
=&J(g) + 2\langle g, \rho\rangle_{\mathcal{H}_K} + J(\rho) \\
=&J(g) + J(\rho) \\
\geq&J(g)
\end{aligned}
$$

with equality iff $\rho(x) = 0$.
Then, we can get the conclusion:

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

with equality iff $\rho(x) = 0$. □

Ex. 5.16

Solution. By the property of RKHS, we can know that $\{\phi_i(x)\}$ is an orthonormal basis. which means:

$$\forall m, k \in \mathbb{N}, \int \phi_m(x)\phi_k(x)\,\mathrm{d}x = \delta(m,k) = \begin{cases} 1 & m = k \\ 0 & m \neq k \end{cases}$$

According to the definition of the Kernel, we have

$$\begin{aligned}
&\sum_{m=1}^{\infty} \gamma_m \phi_m(x)\phi_m(y) = K(x,y) = \sum_{m=1}^{M} h_m(x)h_m(y) \\
\Longrightarrow &\sum_{m=1}^{\infty} \gamma_m \phi_m(x)\phi_m(y)\phi_k(x)\phi_l(y) = \sum_{m=1}^{M} h_m(x)h_m(y)\phi_k(x)\phi_l(y) \\
\Longrightarrow &\int \left(\sum_{m=1}^{\infty} \int \gamma_m \phi_m(x)\phi_m(y)\phi_k(x)\phi_l(y)\mathrm{d}x \right) \mathrm{d}y = \int \left(\sum_{m=1}^{M} \int h_m(x)h_m(y)\phi_k(x)\phi_l(y)\mathrm{d}x \right) \mathrm{d}y \\
\Longrightarrow &\int \left(\sum_{m=1}^{\infty} \gamma_m \phi_m(y)\phi_l(y) \int \phi_m(x)\phi_k(x)\mathrm{d}x \right) \mathrm{d}y = \int \left(\sum_{m=1}^{M} h_m(y)\phi_l(y) \int h_m(x)\phi_k(x)\mathrm{d}x \right) \mathrm{d}y \\
\Longrightarrow &\int \gamma_k \phi_k(y)\phi_l(y)\mathrm{d}y = \int \sum_{m=1}^{M} h_m(y)\phi_l(y)\mathrm{d}y \int h_m(x)\phi_k(x)\mathrm{d}x \\
\Longrightarrow &\gamma_k \delta(k,l) = \sum_{m=1}^{M} \int h_m(y)\phi_l(y)\mathrm{d}y \int h_m(x)\phi_k(x)\mathrm{d}x
\end{aligned} \tag{1}$$

Suppose that G is an $M \times M$ matrix and $G_{km} = \int h_m(x)\phi_k(x)\mathrm{d}x$.
Then, according to (1), we have $\mathbf{G}\mathbf{G}^T = diag(\gamma_1, \gamma_2, \ldots, \gamma_M) = \mathbf{D}_\gamma$.

Assume that $\phi(x) = [\phi_1(x), \phi_2(x), \ldots, \phi_M(x)]^T$, and $h(x) = [h_1(x), h_2(x), \ldots, h_M(x)]^T$
We have

$$
\begin{aligned}
&\sum_{m=1}^{\infty} \gamma_m \phi_m(x)\phi_m(y) = K(x,y) = \sum_{m=1}^{M} h_m(x)h_m(y) \\
\Longrightarrow &\sum_{m=1}^{\infty} \int \gamma_m \phi_m(x)\phi_m(y)\phi_k(y)\mathrm{d}y = \sum_{m=1}^{M} \int h_m(x)h_m(y)\phi_k(y)\mathrm{d}y \\
\Longrightarrow &\sum_{m=1}^{\infty} \gamma_m \phi_m(x)\delta(m,k) = \sum_{m=1}^{M} h_m(x) \int h_m(y)\phi_k(y)\mathrm{d}y \\
\Longrightarrow &\gamma_k \phi_k(x) = \sum_{m=1}^{M} h_m(x)\mathbf{G}_{km} \\
\Longrightarrow &\mathbf{D}_\gamma \phi(x) = \mathbf{G}h(x) \\
\Longrightarrow &h(x) = \mathbf{G}^{-1}\mathbf{D}_\gamma^{\frac{1}{2}}\mathbf{D}_\gamma^{\frac{1}{2}}\phi(x)
\end{aligned}
$$

Let $\mathbf{V} = \mathbf{G}^{-1}\mathbf{D}_\gamma^{\frac{1}{2}}$, we can show that $\mathbf{V}$ is an orthogonal matrix:

$$
\begin{aligned}
\mathbf{V}^T\mathbf{V} =&(\mathbf{G}^{-1}\mathbf{D}_\gamma^{\frac{1}{2}})^T\mathbf{G}^{-1}\mathbf{D}_\gamma^{\frac{1}{2}} \\
=&\mathbf{D}_\gamma^{\frac{1}{2}}(\mathbf{G}^T)^{-1}\mathbf{G}^{-1}\mathbf{D}_\gamma^{\frac{1}{2}} \\
=&\mathbf{D}_\gamma^{\frac{1}{2}}(\mathbf{G}\mathbf{G}^T)^{-1}\mathbf{D}_\gamma^{\frac{1}{2}} \\
=&\mathbf{D}_\gamma^{\frac{1}{2}}\mathbf{D}_\gamma^{-1}\mathbf{D}_\gamma^{\frac{1}{2}} \\
=&\mathbf{I}
\end{aligned}
$$

Therefore, $h(x) = \mathbf{V}\mathbf{D}_\gamma^{\frac{1}{2}}\phi(x)$.
Then, we prove that (5.63) is equivalent to (5.53).

Let $\beta = [\beta_1, \beta_2, \dots, \beta_M]^T$, $c = \mathbf{D}_\gamma^{\frac{1}{2}} \mathbf{V}^T \beta$, we have

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

(b) We can rewrite (5.63) as $\min\limits_{\beta} (y - \mathbf{H}\beta)^T (y - \mathbf{H}\beta) + \lambda \beta^T \beta$ and solve β:

$$\hat{\beta} = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T y$$

As $\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I}$ is positive definite matrix, it is also invertible, and so as $\mathbf{H}\mathbf{H}^T + \lambda \mathbf{I}$.
Then, we notice the following result:

$$\begin{aligned}
&\mathbf{H}^T \mathbf{H} \mathbf{H}^T + \lambda \mathbf{H}^T = \mathbf{H}^T \mathbf{H} \mathbf{H}^T + \lambda \mathbf{H}^T \\
\iff &\mathbf{H}^T (\mathbf{H}\mathbf{H}^T + \lambda \mathbf{I}) = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I}) \mathbf{H}^T \\
\iff &(\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T = \mathbf{H}^T (\mathbf{H}\mathbf{H}^T + \lambda \mathbf{I})^{-1} \\
\Longrightarrow &\mathbf{H} (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T y = \mathbf{H}\mathbf{H}^T (\mathbf{H}\mathbf{H}^T + \lambda \mathbf{I})^{-1} y
\end{aligned}$$

So, we have

$$\begin{aligned}
\hat{\mathbf{f}} =& \mathbf{H}\hat{\beta} \\
=& \mathbf{H} (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T y \\
=& \mathbf{H}\mathbf{H}^T (\mathbf{H}\mathbf{H}^T + \lambda \mathbf{I})^{-1} y \\
=& \mathbf{K} (\mathbf{K} + \lambda \mathbf{I})^{-1} y
\end{aligned}$$

(c) Using the result of (b), we have $\hat{\beta} = \mathbf{H}^T \hat{\alpha}$
Then,

$$\begin{aligned}\hat{f}(x) =& h(x)^T \hat{\beta} \\ =& h(x)^T \mathbf{H}^T \hat{\alpha} \\ =& h(x)^T [h_1(x), h_2(x), \ldots, h_N(x)] \hat{\alpha} \\ =& [K(x, x_1), K(x, x_2), \ldots, K(x, x_N)][\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_N]^T \\ =& \sum_{i=1}^{N} K(x, x_i) \hat{\alpha}_i\end{aligned}$$

(d)When $M < N$,K may be not invertible.
If $\lambda > 0$, since $(\mathbf{K} + \lambda \mathbf{I})^{-1}$ is still invertible, the result of (b) and (c) still hold.
If $\lambda = 0$, since $\mathbf{H}^T \mathbf{H}$ is still invertible, we have

$$\hat{\beta} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T y$$

So,

$$\hat{f}(x) = h(x)^T \hat{\beta} = h(x)^T (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T y$$

$\square$