- 7.1

$$
\begin{aligned}
\sum_{i=1}^{N}\mathrm{Cov}(\hat{y}_i, y_i) &= \mathrm{trace}(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\sigma_\epsilon^2\\
&= \mathrm{trace}((\mathbf{X}^T\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1})\sigma_\epsilon^2\\
&= \mathrm{trace}(I)\sigma_\epsilon^2\\
&= d\sigma_\epsilon^2
\end{aligned}
$$

又因为

$$
E_{\mathbf{y}}(\mathrm{Err_{in}}) = E_{\mathbf{y}}(\overline{\mathrm{err}}) + E(\mathrm{op}) = E_{\mathbf{y}}(\overline{\mathrm{err}}) + \frac{2}{N}\sum_{i=1}^{N}\mathrm{Cov}(\hat{y}_i, y_i)
$$

所以

$$
E_{\mathbf{y}}(\mathrm{Err_{in}}) = E_{\mathbf{y}}(\overline{\mathrm{err}}) + \frac{2}{N}d\sigma_\epsilon^2
$$

也就是式(7.24)。

- 7.3

（1）

$$
\begin{aligned}
\hat{f}^{-i}(x_i) &= x_i^T(\mathbf{X}_{-i}^T\mathbf{X}_{-i} + \lambda\mathbf{\Omega})^{-1}\mathbf{X}_{-i}^T\mathbf{y}_{-i}\\
&= x_i^T(\mathbf{X}^T\mathbf{X} - x_ix_i^T + \lambda\mathbf{\Omega})^{-1}(\mathbf{X}^T\mathbf{y} - x_iy_i)
\end{aligned}
$$

由Woodbury矩阵恒等式

$$
(\mathbf{A} + UCV)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}U(C^{-1} + V\mathbf{A}^{-1}U)^{-1}V\mathbf{A}^{-1}
$$

令$U = x_i, C = 1, V = x_i^T, \mathbf{A} = (\mathbf{X}^T\mathbf{X} + \lambda\Omega)$。

$$
\begin{aligned}
(\mathbf{A} - x_ix_i^T)^{-1} &= \mathbf{A}^{-1} + \mathbf{A}^{-1}x_i(1 - x_i^T\mathbf{A}^{-1}x_i)^{-1}x_i^T\mathbf{A}^{-1}\\
&= \mathbf{A}^{-1} + \frac{\mathbf{A}^{-1}x_ix_i^T\mathbf{A}^{-1}}{1 - x_i^T\mathbf{A}^{-1}x_i}
\end{aligned}
$$

于是

$$
\begin{aligned}
\hat{f}^{-i}(x_i) &= x_i^T\left(\mathbf{A}^{-1} + \frac{\mathbf{A}^{-1}x_ix_i^T\mathbf{A}^{-1}}{1 - x_i^T\mathbf{A}^{-1}x_i}\right)(\mathbf{X}^T\mathbf{y} - x_iy_i)\\
&= \left(x_i^T\mathbf{A}^{-1} + \frac{x_i^T\mathbf{A}^{-1}x_ix_i^T\mathbf{A}^{-1}}{1 - S_{ii}}\right)(\mathbf{X}^T\mathbf{y} - x_iy_i)
\end{aligned}
$$

其中$x_i^T\mathbf{A}^{-1}x_i = x_i^T(\mathbf{X}^T\mathbf{X} + \lambda\Omega)^{-1}x_i = S_{ii}$，

$$
\begin{aligned}
\hat{f}^{-i}(x_i) &= \left(x_i^T\mathbf{A}^{-1} + \frac{S_{ii}x_i^T\mathbf{A}^{-1}}{1 - S_{ii}}\right)(\mathbf{X}^T\mathbf{y} - x_iy_i)\\
&= x_i^T\mathbf{A}^{-1}\mathbf{X}^T\mathbf{y} - x_i^T\mathbf{A}^{-1}x_iy_i + \frac{S_{ii}x_i^T\mathbf{A}^{-1}\mathbf{X}^T\mathbf{y}}{1 - S_{ii}} - \frac{S_{ii}x_i^T\mathbf{A}^{-1}x_iy_i}{1 - S_{ii}}
\end{aligned}
$$

其中$x_i^T\mathbf{A}^{-1}\mathbf{X}^T\mathbf{y} = x_i^T(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{\Omega})^{-1}\mathbf{X}^T\mathbf{y} = \hat{f}(x_i)$。

$$
\begin{aligned}
\hat{f}^{-i}(x_i) &= \hat{f}(x_i) - S_{ii}y_i + \frac{S_{ii}\hat{f}(x_i)}{1 - S_{ii}} - \frac{S_{ii}^2y_i}{1 - S_{ii}}\\
&= \frac{\hat{f}(x_i) - S_{ii}y_i}{1 - S_{ii}}
\end{aligned}
$$

所以

$$y_i - \hat{f}^{-i}(x_i) = \frac{y_i - \hat{f}(x_i)}{1 - S_{ii}}$$

（2）

由于$\mathbf{S}$是实对称矩阵且$\mathbf{S}\mathbf{S} \preceq \mathbf{S}$，所以

$$S_{ii} \geq \sum_k S_{ik}S_{ki} = \sum_k S_{ik}^2 \geq S_{ii}^2$$

有$0 \leq S_{ii} \leq 1$，结合上一问的结论

$$\begin{aligned}\left|y_i - \hat{f}^{-i}(x_i)\right| &= \left|\frac{y_i - \hat{f}(x_i)}{1 - S_{ii}}\right| \\ &= \left|y_i - \hat{f}(x_i)\right| \cdot \frac{1}{|1 - S_{ii}|} \\ &\geq \left|y_i - \hat{f}(x_i)\right|\end{aligned}$$

（3）

只需要$\mathbf{S}$与$\mathbf{y}$无关，第一问中的结论就依然成立。

把平滑矩阵的一般形式写成$\mathbf{S} = \mathbf{N}(\mathbf{N}^T\mathbf{N} + \lambda\mathbf{\Omega})^{-1}\mathbf{N}^T$。$\mathbf{N}$是一个与$\mathbf{X}$有关的同尺寸矩阵，且$\mathbf{N}$的第$i$列只与$x_i$有关。那么我们可以将第一问中所有的$\mathbf{X}$用$\mathbf{N}$替换，$x_i$用$N_i$替换，第一问中的推导依然成立。

- 7.4

$$\begin{aligned}E(\text{op}) &= E[\text{Err}_{\text{in}} - \overline{\text{err}}] \\ &= E\left[\frac{1}{N}\sum_{i=1}^N E_{Y^0}(Y^0 - \hat{f}(x_i))^2 - \frac{1}{N}\sum_{i=1}^N (y_i - \hat{f}(x_i))^2\right] \\ &= \frac{1}{N}E\left[\sum_{i=1}^N -y_i^2 - y_i E_{Y^0}Y^0 + 2\hat{f}(x_i)y_i + E_{Y^0}Y^0 y_i + E_{Y^0}{Y^0}^2 - 2E_{Y^0}Y^0\hat{f}(x_i)\right]\end{aligned}$$

上式中

$$E\sum_{i=1}^N -y_i^2 E_{Y^0}{Y^0}^2 + E_{Y^0}{Y^0}^2 = \sum_{i=1}^N E(E_{Y^0}{Y^0}^2 - y_i^2 E_{Y^0}{Y^0}^2) = 0$$

$$E\sum_{i=1}^N -y_i E_{Y^0}Y^0 + E_{Y^0}Y^0 y_i = 0$$

所以

$$\begin{aligned}E(\text{op}) &= \frac{1}{N}E\sum_{i=1}^N 2\hat{f}(x_i)y_i - 2E_{Y^0}Y^0\hat{f}(x_i) \\ &= \frac{2}{N}\sum_{i=1}^N \text{Cov}(\hat{f}(x_i), y_i)\end{aligned}$$

- 7.5

$$\begin{aligned}\sum_{i=1}^N \text{Cov}(\hat{y}_i, y_i) &= \text{trace}(\text{Cov}(\mathbf{S}\mathbf{y}, \mathbf{y})) \\ &= \text{trace}(\mathbf{S}\text{Var}(\mathbf{y})) \\ &= \text{trace}(\mathbf{S})\sigma_\epsilon^2\end{aligned}$$