第六次作业

Ext 6.2

Ex. 6.2 Show that $\sum_{i=1}^N (x_i - x_0) l_i(x_0) = 0$ for local linear regression. Define $b_j(x_0) = \sum_{i=1}^N (x_i - x_0)^j l_i(x_0)$. Show that $b_0(x_0) = 1$ for local polynomial regression of any degree (including local constants). Show that $b_j(x_0) = 0$ for all $j \in \{1, 2, \ldots, k\}$ for local polynomial regression of degree k. What are the implications of this on the bias?

Define the vector-valued function $b(x)^T = \left(1, x, x^2, \ldots, x^k\right)$ for $k \geq 0$. Let B be the $N \times (k+1)$ regression matrix with i th row $b\left(x_i\right)^T$, and $\mathbf{W}\left(x_0\right)$ the $N \times N$ diagonal matrix with i th diagonal element $K_\lambda\left(x_0, x_i\right)$. Then we have：

$$b\left(x_0\right)^T = b\left(x_0\right)^T\left(\mathbf{B}^T \mathbf{W}\left(x_0\right) \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{W}\left(x_0\right) \mathbf{B}.$$

And we can rewrite the sum of $l_i(x_0)$

$$\begin{aligned}
1 &= b\left(x_0\right)^T\left(\mathbf{B}^T \mathbf{W}\left(x_0\right) \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{W}\left(x_0\right) \mathbf{1} = \sum_{i=1}^N l_i\left(x_0\right) \\
x_0 &= b\left(x_0\right)^T\left(\mathbf{B}^T \mathbf{W}\left(x_0\right) \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{W}\left(x_0\right) \mathbf{B}_2 = \sum_{i=1}^N l_i\left(x_0\right) x_i \\
x_0^k &= b\left(x_0\right)^T\left(\mathbf{B}^T \mathbf{W}\left(x_0\right) \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{W}\left(x_0\right) \mathbf{B}_{k+1} = \sum_{i=1}^N l_i\left(x_0\right) x_i^k
\end{aligned}$$

where $\mathbf{B}_i$ is the i-th column of $\mathbf{B}$. Therefore we have $b_0\left(x_0\right) = \sum i = 1^N l_i\left(x_0\right) = 1$ and

$$b_1\left(x_0\right) = \sum i = 1^N\left(x_i - x_0\right) l_i\left(x_0\right) = \sum_{i=1}^N l_i\left(x_0\right) x_i - x_0 \sum_{i=1}^N l_i\left(x_0\right) = x_0 - x_0 \cdot 1 = 0.$$

For $j \geq 2$, we have

$$
\begin{aligned}
b_j\left(x_0\right) &= \sum_{i=1}^{N}\left(x_i-x_0\right)^j l_i\left(x_0\right) \\
&= \sum_{i=1}^{N}\left(\sum_{b=0}^{j} C_j^b(-1)^b x_i^{j-b} x_0^b\right) l_i\left(x_0\right) \\
&= \sum_{b=0}^{j} C_j^b(-1)^b x_0^b\left(\sum_{i=1}^{N} l_i\left(x_0\right) x_i^{j-b}\right) \\
&= \sum^{j} C_j^b(-1)^b x_0^b x_0^{j-b} \\
&= \sum_{b=0}^{j} C_j^b(-1)^b x_0^j \\
&= (1-1)^j x_0^j \\
&= 0.
\end{aligned}
$$

Use the taylor expansion we get that the reminder only contains the (k+1)-th and higher-order deriviatives of f.

Ext 6.7

Ex. 6.7 Derive an expression for the leave-one-out cross-validated residual sum-of-squares for local polynomial regression.

Note that local regression smoothers are linear estimators, and we can write

$$
\hat{\mathbf{f}}=\mathbf{S}_{\lambda \mathbf{y}}
$$

where $\{\mathbf{S} \lambda\} i j=l_i\left(x_j\right)$ for $l_i(x)$ defined by (6.8) in the text. Then we know

$$
y_i-\hat{f}^{-i}\left(x_i\right)=\frac{y_i-\hat{f}\left(x_i\right)}{1-\{\mathbf{S} \lambda\} i i}
$$

Ext 6.10

Ex. 6.10 Suppose we have N samples generated from the model $y_i = f(x_i) + \varepsilon_i$, with ε_i independent and identically distributed with mean zero and variance σ^2, the x_i assumed fixed (non random). We estimate f using a linear smoother (local regression, smoothing spline, etc.) with smoothing parameter λ. Thus the vector of fitted values is given by $\hat{\mathbf{f}} = \mathbf{S}_\lambda \mathbf{y}$. Consider the *in-sample* prediction error

$$\mathrm{PE}(\lambda) = \mathrm{E}\frac{1}{N}\sum_{i=1}^{N}(y_i^* - \hat{f}_\lambda(x_i))^2 \tag{6.34}$$

for predicting new responses at the N input values. Show that the average squared residual on the training data, $\mathrm{ASR}(\lambda)$, is a biased estimate (optimistic) for $\mathrm{PE}(\lambda)$, while

$$C_\lambda = \mathrm{ASR}(\lambda) + \frac{2\sigma^2}{N}\mathrm{trace}(\mathbf{S}_\lambda) \tag{6.35}$$

is unbiased.

From the view of chapter7, we can know that $PE(\lambda) = Err_{in}, ASR(\lambda) = \overline{err}$.

So we only need to prove that:

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

So.

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