Homework of Chapter 6

Chen Cheng, 1130339005

January 11, 2015

Ex. 6.2

Solution. According to (6.7),we can find that $\sum_{i=1}^{N}(x_i - x_0)l_i(x_0)$ is the solution of

$$\min_{\alpha(x_0),\beta(x_0)} \sum_{i=1}^{N} K_\lambda(x_0, x_i)[(x_i - x_0) - \alpha(x_0) - \beta(x_0)x_i]^2$$

which is not less than 0.
We set $\hat{\alpha(x_0)} = -x_0$ and $\hat{\beta(x_0)} = 1$, and the above expression gets the minimum value. Then we have $\hat{f}(x_0) = \hat{\alpha(x_0)} + \hat{\beta(x_0)}x_0 = 0$, namely, $\sum_{i=1}^{N}(x_i - x_0)l_i(x_0) = 0$.
Similarly, we note that $b_0(x_0) = \sum_{i=1}^{N}(x_i - x_0)^0 l_i(x_0) = \sum_{i=1}^{N} l_i(x_0)$ is the solution of

$$\min_{\alpha(x_0),\beta_j(x_0),j=1,\ldots,d} \sum_{i=1}^{N} K_\lambda(x_0, x_i)\left[1 - \alpha(x_0) - \sum_{j=1}^{d}\beta_j(x_0)x_i^j\right]^2$$

We set $\hat{\alpha(x_0)} = 1$ and $\hat{\beta(x_0)} = 0$, and the above expression gets the minimum value. Then we have $\hat{f(x_0)} = 1$. Namely, $b_0(x_0) = \sum_{i=1}^{N} l_i(x_0) = 1$.
$b_k(x_0) = \sum_{i=1}^{N}(x_i - x_0)^k l_i(x_0)$ is the solution of

$$\min_{\alpha(x_0),\beta_j(x_0),j=1,\ldots,d} \sum_{i=1}^{N} K_\lambda(x_0, x_i)\left[(x_i - x_0)^k - \alpha(x_0) - \sum_{j=1}^{d}\beta_j(x_0)x_i^j\right]^2$$

We set $\alpha(x_0)\hat{} = x_0^k$ and

$$\hat{\beta}_j(x_0) = \begin{cases} C_k^j(-1)^j x_0^{k-j} & j = 1, \ldots, k \\ 0 & j = k+1, \ldots, d \end{cases}$$

Hence, we have

$$\begin{aligned} \hat{f}(x_0) =& \hat{\alpha}(x_0) + \sum_{j=1}^{d} \hat{\beta}_j(x_0) x_0^j \\ =& \sum_{j=0}^{k} C_k^j (-1)^j x_0^{k-j} x_0^j \\ =& (x_0 - x_0)^k \\ =& 0 \end{aligned}$$

So, $b_j(x_0) = \hat{f}(x_0) = 0$
According to (6.10), we have

$$E(\hat{f}(x_0)) = f(x_0) \sum_{i=1}^{N} l_i(x_0) + \sum_{j=1}^{k} \frac{f^{(j)}(x_0)}{j!} \sum_{i=1}^{N} (x_i - x_0)^j l_i(x_0) + R$$

where the remainder term R involves $k+1$ and higher-order derivatives of f. As $\sum_{i=1}^{N} l_i(x_0) = 1$,$\sum_{i=1}^{N}(x_i - x_0)^j l_i(x_0) = 0, j = 1, \ldots, k$, we have that the bias for local polynomial regression only related to $k+1$ or higher order terms in the expansion of f. $\square$

Ex. 6.3

Solution.

$$
\begin{aligned}
\| l(x_0) \|^2 =& l(x_0)^T l(x_0) \\
=& b(x_0)^T (B^T W(x_0) B)^{-1} B^T W(x_0) W(x_0) B (B^T W(x_0) B)^{-1} b(x_0)
\end{aligned}
$$

Consider the problem:

$$
\min_{\alpha(x_0), \beta(x_0)} l_p(\beta, x_0) = \sum_{i=1}^{N} K_\lambda(x_0, x_i)[y_i - [1, x_i^T]\beta(x_0)]^2
$$

Assume $\hat{\beta}_p(x_0) = var \min_\beta l_p(\beta, x_0)$.
According to minimization problem, we have

$$
l_p(\hat{\beta}_p(x_0), x_0) \geq l_{p+1}(\hat{\beta}_{p+1}(x_0), x_0)
$$

On the other hand,

$$
\begin{aligned}
l_p(\hat{\beta}_p(x_0), x_0) =& [Y - B_p \hat{\beta}_p(x_0)]^T W(x_0) [Y - B_p \hat{\beta}_p(x_0)] \\
=& Y^T W(x_0) Y - 2 Y^T W(x_0) B_p \hat{\beta}_p(x_0) + [B_P \hat{\beta}_p(x_0)]^T W(x_0) B_p \hat{\beta}_p(x_0) \\
=& Y^T W(x_0) Y - Y^T W(x_0) B_p (B_p^T W(x_0) B_p)^{-1} B_p^T W(x_0) Y
\end{aligned}
$$

Combining the above equation with the relation:

$$
l_p(\hat{\beta}_p(x_0), x_0) \geq l_{p+1}(\hat{\beta}_{p+1}(x_0), x_0)
$$

We have the following inequality:

$$
\begin{aligned}
& Y^T W(x_0) B_p (B_p^T W(x_0) B_p)^{-1} B_p^T W(x_0) Y \\
\leq & Y^T W(x_0) B_{p+1} (B_{p+1}^T W(x_0) B_{p+1})^{-1} B_{p+1}^T W(x_0) Y
\end{aligned}
$$

which holds for every vector Y, in particular

$$
\begin{aligned}
Y =& W(x_0)^{\frac{1}{2}} B (B^T W(x_0) B)^{-1} b(x_0) \\
Y^T =& b(x_0)^T (B^T W(x_0) B)^{-1} B^T W(x_0)^{\frac{1}{2}}
\end{aligned}
$$

So,

$$\begin{aligned}
&Y^T W(x_0) B_p (B_p^T W(x_0) B_p)^{-1} B_p^T W(x_0) Y \\
=&b(x_0)^T (B^T W(x_0) B)^{-1} B^T W(x_0)^{\frac{1}{2}} \times \\
&W(x_0) B_p (B_p^T W(x_0) B_p)^{-1} B_p^T W(x_0) \times \\
&W(x_0)^{\frac{1}{2}} B (B^T W(x_0) B)^{-1} b(x_0) \\
=&b(x_0)^T (B^T W(x_0) B)^{-1} B^T \times \\
&W(x_0) B_p (B_p^T W(x_0) B_p)^{-1} B_p^T W(x_0) \times \\
&W(x_0)^{\frac{1}{2}} W(x_0)^{\frac{1}{2}} B (B^T W(x_0) B)^{-1} b(x_0) \\
=&b(x_0)^T (B^T W(x_0) B)^{-1} B^T W(x_0) \times \\
&W(x_0) B (B^T W(x_0) B)^{-1} b(x_0) \\
=&\parallel l_p(x_0) \parallel^2
\end{aligned}$$

So

$$\parallel l_p(x_0) \parallel^2 \leq \parallel l_{p+1}(x_0) \parallel^2$$

which means that $\parallel l(x) \parallel^2$ increase with the degree in the local polynomial. $\square$

Ex. 6.4

Solution. $D = \sqrt{(x - x_0)^T \mathbf{\Sigma}^{-1} (x - x_0)}$ is called the *Mahalanobis distance*. The Mahalanobis distance takes into account the correlations of different data components. When $\mathbf{A} = \mathbf{I}$, the Mahalanobis distance is the same with Euclidean distance, which means all the components of the data are independent.

(a) To downweights high-frequency components x_i, x_j in the distance metric, we can just decrease $Cov(x_i, x_j)$.

(b) To ignore the high-frequency completely, we can just set $Cov(x_i, x_j) = 0$. $\square$