Homework of Chapter 2

Cheng Chen, 1130339005

January 8, 2014

Ex. 2.4

Solution. Since $X \sim N(\mathbf{0}, \boldsymbol{I}_p)$,
We have

$$Z = a^T X = \frac{x_0^T}{\|x\|} X \sim N(0, \frac{x_0^T}{\|x\|} \boldsymbol{I}_p \frac{x_0}{\|x\|}) = N(0,1)$$

Therefore, z_i are distributed $N(0,1)$. $\square$

Ex. 2.7

Solution.

(a) We define $\boldsymbol{X}$ as textbook did:

$$\boldsymbol{X}_{N\times(p+1)} = \begin{pmatrix} 1 & \boldsymbol{x}_1^T \\ \vdots & \vdots \\ 1 & \boldsymbol{x}_N^T \end{pmatrix}$$

The linear regression is $\hat{f}(\boldsymbol{x}_0) = \begin{pmatrix} 1 & \boldsymbol{x}_0^T \end{pmatrix} \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$.
So,we have

$$\hat{f}(\boldsymbol{x}_0) = \begin{pmatrix} 1 & \boldsymbol{x}_0^T \end{pmatrix} (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

In terms of the notation of the question,

$$\ell_i(\boldsymbol{x}_0; \mathcal{X}) = \begin{pmatrix} 1 & \boldsymbol{x}_0^T \end{pmatrix} (\boldsymbol{X}^T\boldsymbol{X})^{-1} \begin{pmatrix} 1 \\ \boldsymbol{x}_i \end{pmatrix}$$

Consider the k-nearest-neighbour regression,

$$\ell_i(\boldsymbol{x}_0;\mathcal{X}) = \begin{cases} \frac{1}{k} & \boldsymbol{x}_0 \in S \\ 0 & \boldsymbol{x}_0 \notin S \end{cases}$$

where S is the set of k-nearest-neighbour points.
(b)

$$\begin{aligned}
&E_{\mathcal{Y}|\mathcal{X}}(f(\boldsymbol{x_0}) - \hat{f}(\boldsymbol{x}_0))^2 \\
=&f^2(\boldsymbol{x_0}) + E_{\mathcal{Y}|\mathcal{X}}(\hat{f}^2(\boldsymbol{x}_0)) - 2f(\boldsymbol{x_0})E_{\mathcal{Y}|\mathcal{X}}(\hat{f}(\boldsymbol{x}_0)) \\
=&(f(\boldsymbol{x_0}) - E_{\mathcal{Y}|\mathcal{X}}(\hat{f}(\boldsymbol{x}_0)))^2 + (E_{\mathcal{Y}|\mathcal{X}}(\hat{f}^2(\boldsymbol{x}_0)) - E_{\mathcal{Y}|\mathcal{X}}(\hat{f}(\boldsymbol{x}_0))^2) \\
=&(Bias_{\mathcal{Y}|\mathcal{X}})^2 + Var_{\mathcal{Y}|\mathcal{X}}(\hat{f}(\boldsymbol{x}_0))
\end{aligned}$$

(c)

$$\begin{aligned}
&E_{\mathcal{Y},\mathcal{X}}(f(\boldsymbol{x_0}) - \hat{f}(\boldsymbol{x}_0))^2 \\
=&(f(\boldsymbol{x_0}) - E_{\mathcal{Y},\mathcal{X}}(\hat{f}(\boldsymbol{x}_0)))^2 + (E_{\mathcal{Y},\mathcal{X}}(\hat{f}^2(\boldsymbol{x}_0)) - E_{\mathcal{Y},\mathcal{X}}(\hat{f}(\boldsymbol{x}_0))^2) \\
=&(Bias_{\mathcal{Y},\mathcal{X}})^2 + Var_{\mathcal{Y},\mathcal{X}}(\hat{f}(\boldsymbol{x}_0))
\end{aligned}$$

(d) First of all,Since

$$E_{\mathcal{X}}(E_{\mathcal{Y}|\mathcal{X}}(f(\boldsymbol{x_0}) - \hat{f}(\boldsymbol{x}_0))^2) = E_{\mathcal{Y},\mathcal{X}}(f(\boldsymbol{x_0}) - \hat{f}(\boldsymbol{x}_0))^2$$

We have

$$E_{\mathcal{X}}(Bias^2_{\mathcal{Y}|\mathcal{X}}) + E_{\mathcal{X}}(Var_{\mathcal{Y}|\mathcal{X}}(\hat{f}(\boldsymbol{x}_0))) = (Bias_{\mathcal{Y},\mathcal{X}})^2 + Var_{\mathcal{Y},\mathcal{X}}(\hat{f}(\boldsymbol{x}_0))$$

Secondly,

$$\begin{aligned}
&E_{\mathcal{Y}|\mathcal{X}}(\hat{f}(\boldsymbol{x}_0)) \\
=&E_{\mathcal{Y}|\mathcal{X}}\left(\sum_{i=1}^{N} \ell_i(\boldsymbol{x}_0;\mathcal{X})y_i\right) \\
=&E_{\mathcal{Y}|\mathcal{X}}\left(\sum_{i=1}^{N} \ell_i(\boldsymbol{x}_0;\mathcal{X})(f(\boldsymbol{x}_i) + \varepsilon_i)\right) \\
=&\sum_{i=1}^{N} \ell_i(\boldsymbol{x}_0;\mathcal{X})f(\boldsymbol{x}_i)
\end{aligned}$$

So,

$$
\begin{aligned}
&(Bias_{\mathcal{Y},\mathcal{X}})^2\\
=&(f(\boldsymbol{x_0}) - E_{\mathcal{Y},\mathcal{X}}(\hat{f}(\boldsymbol{x}_0)))^2\\
=&(f(\boldsymbol{x_0}) - E_{\mathcal{X}}(E_{\mathcal{Y}|\mathcal{X}}(\hat{f}(\boldsymbol{x}_0))))^2\\
=&\left(f(\boldsymbol{x_0}) - E_{\mathcal{X}}\left(\sum_{i=1}^{N} \ell_i(\boldsymbol{x}_0;\mathcal{X})f(\boldsymbol{x}_i)\right)\right)^2\\
=&\left(E_{\mathcal{X}}\left(f(\boldsymbol{x_0}) - \sum_{i=1}^{N} \ell_i(\boldsymbol{x}_0;\mathcal{X})f(\boldsymbol{x}_i)\right)\right)^2\\
\leq&E_{\mathcal{X}}\left(f(\boldsymbol{x_0}) - \sum_{i=1}^{N} \ell_i(\boldsymbol{x}_0;\mathcal{X})f(\boldsymbol{x}_i)\right)^2\\
=&E_{\mathcal{X}}((Bias_{\mathcal{Y}|\mathcal{X}})^2)
\end{aligned}
$$

Since

$$E_{\mathcal{X}}((Bias_{\mathcal{Y}|\mathcal{X}})^2) + E_{\mathcal{X}}(Var_{\mathcal{Y}|\mathcal{X}}(\hat{f}(\boldsymbol{x}_0))) = (Bias_{\mathcal{Y},\mathcal{X}})^2 + Var_{\mathcal{Y},\mathcal{X}}(\hat{f}(\boldsymbol{x}_0))$$

We can achieve the relationship between the squared biases and variances as follows:

$$(Bias_{\mathcal{Y},\mathcal{X}})^2 \leq E_{\mathcal{X}}((Bias_{\mathcal{Y}|\mathcal{X}})^2)$$

and,

$$Var_{\mathcal{Y},\mathcal{X}} \geq E_{\mathcal{X}}(Var_{\mathcal{Y}|\mathcal{X}})$$

$\square$