Homework 6

Runmin Lu

October 12, 2021

Exercise 3.4

(a)

$$\begin{aligned}
\hat{\mathbf{y}} &= \mathbf{X}\mathbf{w}_{\text{lin}} \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\mathbf{w}^* + \boldsymbol{\epsilon}) \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\mathbf{w}^* + \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \\
&= \mathbf{X}\mathbf{w}^* + \mathbf{H}\boldsymbol{\epsilon}
\end{aligned}$$

(b)

$$\begin{aligned}
\hat{\mathbf{y}} - \mathbf{y} &= \mathbf{X}\mathbf{w}^* + \mathbf{H}\boldsymbol{\epsilon} - (\mathbf{X}\mathbf{w}^* + \boldsymbol{\epsilon}) \\
&= \mathbf{H}\boldsymbol{\epsilon} - \boldsymbol{\epsilon} \\
&= (\mathbf{H} - \mathbf{I})\boldsymbol{\epsilon}
\end{aligned}$$

(c)

$$
\begin{aligned}
E_{\text{in}}(\mathbf{w}_{\text{lin}}) &= \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \\
&= \frac{1}{N}||\hat{\mathbf{y}} - \mathbf{y}||^2 \\
&= \frac{1}{N}||(\mathbf{H} - \mathbf{I})\boldsymbol{\epsilon}||^2 \\
&= \frac{1}{N}||(\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon}||^2 \\
&= \frac{1}{N}\boldsymbol{\epsilon}^T(\mathbf{I} - \mathbf{H})^T(\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon} \\
&= \frac{1}{N}\boldsymbol{\epsilon}^T(\mathbf{I} - \mathbf{H})^2\boldsymbol{\epsilon} \\
&= \frac{1}{N}\boldsymbol{\epsilon}^T(\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon}
\end{aligned}
$$

(d)

$$
\begin{aligned}
\mathbb{E}_{\mathcal{D}}[E_{\text{in}}(\mathbf{w}_{\text{lin}})] &= \mathbb{E}_{\mathcal{D}}[\frac{1}{N}\boldsymbol{\epsilon}^T(\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon}] \\
&= \frac{1}{N}\mathbb{E}_{\mathcal{D}}[\boldsymbol{\epsilon}^T(\mathbf{I} - \mathbf{H})\boldsymbol{\epsilon}] \\
&= \frac{1}{N}\mathbb{E}_{\mathcal{D}}[\boldsymbol{\epsilon}^T\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^T\mathbf{H}\boldsymbol{\epsilon}] \\
&= \frac{1}{N}(\mathbb{E}_{\mathcal{D}}[||\boldsymbol{\epsilon}||^2] - \mathbb{E}_{\mathcal{D}}[\boldsymbol{\epsilon}^T\mathbf{H}\boldsymbol{\epsilon}]) \\
\mathbb{E}_{\mathcal{D}}[||\boldsymbol{\epsilon}||^2] &= \sum_{i=1}^{N}\mathbb{E}_{\mathcal{D}}[\epsilon_i^2] \\
&= N\sigma^2 \\
\boldsymbol{\epsilon}^T\mathbf{H}\boldsymbol{\epsilon} &= \sum_{i=1}^{N}\sum_{j=1}^{N}H_{ij}\epsilon_i\epsilon_j \\
\mathbb{E}_{\mathcal{D}}[\epsilon_i\epsilon_j] &= \begin{cases} \sigma^2 & i = j \\ 0 & i \neq j \end{cases} \\
\mathbb{E}_{\mathcal{D}}[\boldsymbol{\epsilon}^T\mathbf{H}\boldsymbol{\epsilon}] &= \sum_{i=1}^{N}H_{ii}\sigma^2 \\
&= \sigma^2\text{trace}(\mathbf{H}) \\
&= \sigma^2(d+1) \\
\mathbb{E}_{\mathcal{D}}[E_{\text{in}}(\mathbf{w}_{\text{lin}})] &= \frac{1}{N}(N\sigma^2 - \sigma^2(d+1)) \\
&= \sigma^2(1 - \frac{d+1}{N})
\end{aligned}
$$

(e)

$$
\begin{aligned}
\mathbf{y}' &= \mathbf{X}\mathbf{w}^* + \boldsymbol{\epsilon}' \\
\hat{\mathbf{y}} - \mathbf{y}' &= \mathbf{X}\mathbf{w}^* + \mathbf{H}\boldsymbol{\epsilon} - (\mathbf{X}\mathbf{w}^* + \boldsymbol{\epsilon}') \\
&= \mathbf{H}\boldsymbol{\epsilon} - \boldsymbol{\epsilon}' \\
E_{\text{test}}(\mathbf{w}_{\text{lin}}) &= \frac{1}{N}||\hat{\mathbf{y}} - \mathbf{y}'||^2 \\
&= \frac{1}{N}||\mathbf{H}\boldsymbol{\epsilon} - \boldsymbol{\epsilon}'||^2 \\
&= \frac{1}{N}(\boldsymbol{\epsilon}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\epsilon} - 2\boldsymbol{\epsilon}'^T\mathbf{H}\boldsymbol{\epsilon} + ||\boldsymbol{\epsilon}'||^2) \\
&= \frac{1}{N}(\boldsymbol{\epsilon}^T\mathbf{H}\boldsymbol{\epsilon} - 2\boldsymbol{\epsilon}'^T\mathbf{H}\boldsymbol{\epsilon} + ||\boldsymbol{\epsilon}'||^2) \qquad \text{(from Exercise 3.3 (b))} \\
\mathbb{E}_{\mathcal{D},\boldsymbol{\epsilon}'}[E_{\text{test}}(\mathbf{w}_{\text{lin}})] &= \frac{1}{N}(\mathbb{E}_{\mathcal{D},\boldsymbol{\epsilon}'}[\boldsymbol{\epsilon}^T\mathbf{H}\boldsymbol{\epsilon}] - \mathbb{E}_{\mathcal{D},\boldsymbol{\epsilon}'}[2\boldsymbol{\epsilon}'^T\mathbf{H}\boldsymbol{\epsilon}] + \mathbb{E}_{\mathcal{D},\boldsymbol{\epsilon}'}[||\boldsymbol{\epsilon}'||^2]) \\
\boldsymbol{\epsilon}'^T\mathbf{H}\boldsymbol{\epsilon} &= \sum_{i=1}^{N}\sum_{j=1}^{N} H_{ij}\epsilon'_i\epsilon_j \\
\mathbb{E}_{\mathcal{D},\boldsymbol{\epsilon}'}[\epsilon'_i\epsilon_j] &= 0 \\
\mathbb{E}_{\mathcal{D},\boldsymbol{\epsilon}'}[\boldsymbol{\epsilon}'^T\mathbf{H}\boldsymbol{\epsilon}] &= 0 \\
\mathbb{E}_{\mathcal{D},\boldsymbol{\epsilon}'}[E_{\text{test}}(\mathbf{w}_{\text{lin}})] &= \frac{1}{N}(\sigma^2(d+1) - 0 + N\sigma^2) \\
&= \sigma^2(1 + \frac{d+1}{N})
\end{aligned}
$$