Assignment 6

Darwin Ding

October 12, 2016

Exercise 3.4

(a)

$$
\begin{aligned}
\hat{y} &= Hy \\
&= H(w^{*T}x + \epsilon) \\
&= H(Xw^* + \epsilon) \\
&= HXw^* + H\epsilon \\
&= X(X^TX)^{-1}X^TXw^* + H\epsilon \\
&= XIw^* + H\epsilon \\
&= \boldsymbol{Xw^* + H\epsilon}
\end{aligned}
$$

$w^{*T}x_n + \epsilon_n$ returns the specific y value for a single x vector, and is the nth row in $Xw^* + \epsilon$. Additionally, moving from step 5 to step 6 in the above calculation is legal due to matrix chain multiplication and a matrix being multiplied by its inverse being equal to the identity matrix.

(b)

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

(c) We can adapt the given E_{in} formula to use the full matrices:

$$\begin{aligned} E_{in}(w_{lin}) &= \frac{1}{N}(\hat{y}-y)^T(\hat{y}-y) \\ &= \frac{1}{N}((H-I)\epsilon)^T((H-I)\epsilon) \\ &= \frac{1}{N}\epsilon^T(H-I)^T(H-I)\epsilon \end{aligned}$$

From Exercise 3.3c we know that $(I-H)^T(I-H) = (I-H)$. Since $(I-H) = -1(H-I)$, we can simplify further:

$$\begin{aligned} &\frac{1}{N}\epsilon^T(-1)^2(I-H)^T(I-H)\epsilon \\ &= \frac{1}{N}\epsilon^T(I-H)\epsilon \\ &= \frac{1}{N}\epsilon^T\epsilon - \frac{1}{N}\epsilon^T H\epsilon \end{aligned}$$

(d)

$$\begin{aligned} E_D[E_{in}(w_{lin})] &= E_D[\frac{1}{N}\epsilon^T\epsilon - \frac{1}{N}\epsilon^T H\epsilon] \\ &= E_D[\frac{1}{N}\epsilon^T\epsilon] - E_D[\frac{1}{N}\epsilon^T H\epsilon] \end{aligned}$$

We can reason through the first half of this equation.

$$\begin{aligned} &E_D[\frac{1}{N}\epsilon^T\epsilon] \\ &= \frac{1}{N}E_D[\epsilon^T\epsilon] \end{aligned}$$

Note that $\epsilon^T\epsilon$ is a single value that is the sum of all of the individual noise components squared. Since the variance of each noise component is σ^2 with mean 0 and there are N such noise components:

$$\frac{1}{N}E_D[\epsilon^T\epsilon] = \frac{1}{N} * N\sigma^2 = \sigma^2$$

For the second component, it is helpful to try to visualize what the matrix multiplication will look like and go from there. ϵ^T is a 1 x N matrix, H is an N x N matrix and ϵ is an N x 1 matrix. When we perform the first operation, $\epsilon^T * H$, we essentially end up with a 1 x N matrix as follows:

$$\begin{bmatrix} \epsilon \cdot H_0 & \epsilon \cdot H_1 & \epsilon \cdot H_2 & \dots & \epsilon \cdot H_{N-1} \end{bmatrix}$$

... where H_0 is the first column of H, H_1 is the second column and so forth until H_{N-1}.

This matrix is then multiplied by ϵ, the N x 1 matrix, giving us a final 1 x 1 value:

$$\begin{bmatrix} \epsilon_0 * (\epsilon \cdot H_0) + \epsilon_1 * (\epsilon \cdot H_1) + \dots + \epsilon_{N-1} * (\epsilon \cdot H_{N-1}) \end{bmatrix}$$

We can expand the dot products and factor the ϵ values in.

$$\begin{aligned} &\epsilon_0 * (\epsilon \cdot H_0) + \epsilon_1 * (\epsilon \cdot H_1) + \dots + \epsilon_{N-1} * (\epsilon \cdot H_{N-1}) \\ &= \epsilon_0(\epsilon_0 * H_{0,0} + \epsilon_1 * H_{0,1} + \dots) + \dots + \epsilon_{N-1}(\epsilon_0 * H_{N-1,0} + \epsilon_1 * H_{N-1,1} + \dots) \\ &= \epsilon_0 * \epsilon_0 * H_{0,0} + \epsilon_0 * \epsilon_1 * H_{0,1} + \dots + \epsilon_1 * \epsilon_0 * H_{1,0} + \epsilon_1 * \epsilon_1 * H_{1,1} + \dots \end{aligned}$$

This may look like a random, long and confusing combination of ϵ values and random H values, but there is a pattern! This can all be summarized into summations:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \epsilon_i \epsilon_j H_{i,j}$$

But since we've figured out that $trace(H) = d + 1$ from Exercise 3.3d, it will help to extract the diagonal values:

$$\sum_{i=0}^{N-1} \epsilon_i^2 H_{i,i} + \sum_{i,j=0; i \neq j}^{N-1} \epsilon_i \epsilon_j H_{i,j}$$

From here, we can figure out the expected value with respect to D. First, note that $\sum_{i=0}^{N-1} H_{i,i} = trace(H) = d + 1$. Secondly, note that while the expected value of $\epsilon_N^2 = \sigma^2$ for all N, the expected value of

$\epsilon_N = 0$ for all N. Thus, we can simplify the expression:

$$E[\sum_{i=0}^{N-1} \epsilon_i^2 H_{i,i} + \sum_{i,j=0;i\neq j}^{N-1} \epsilon_i \epsilon_j H_{i,j}]$$
$$= E[\sum_{i=0}^{N-1} \epsilon_i^2 H_{i,i}] + E[\sum_{i,j=0;i\neq j}^{N-1} \epsilon_i \epsilon_j H_{i,j}]$$
$$= \sigma^2(d+1) + 0 = \sigma^2(d+1)$$

So this is the expected value of the matrix multiplication. There is an additional $\frac{1}{N}$ term that needs to be tacked on, but now we can finally combine everything:

$$E_D[\frac{1}{N}\epsilon^T\epsilon] - E_D[\frac{1}{N}\epsilon^T H\epsilon]$$
$$= \sigma^2 - \frac{1}{N}\sigma^2(d+1)$$
$$= \sigma^2(1 - \frac{d+1}{N})$$

Problem 3.1a: 1137 iterations, .1195 slope + 15.274 Problem 3.1b: really quick, .066376x + 16.3346