COL341 - HOMEWORK-1

Rajat Bhardwaj

February 2023

1 Question 1

1.1 A. To prove that H is symmetric

$$H = X(X^TX)^{-1}X^T$$

We have to prove that $H^T = H$
Taking transpose

$$H^T = (X(X^TX)^{-1}X^T)^T$$

Now we know that $(ABC)^T = C^TB^TA^T$
Thus,

$$H^T = (X^T)^T((X^TX)^{-1})^TX^T$$

Now we know that $(A^T)^T = A$ and $(A^{-1})^T = (A^T)^{-1}$
Therefore,

$$H^T = X((X^TX)^T)^{-1}X^T$$
$$H^T = X((X^TX))^{-1}X^T$$
$$H^T = H$$

Which means H is symmetric

1.2 B. To show that $H^k = H$ for $k > 0$

$$H = X(X^TX)^{-1}X^T$$

Let us find H^2

$$H^2 = (X(X^TX)^{-1}X^T)^2$$
$$H^2 = (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T)$$
$$H^2 = (X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T)$$

Cancelling out $(X^TX)^{-1}$ and (X^TX) we get

$$H^2 = X(X^TX)^{-1}X^T$$

Thus

$$H^2 = H$$

Now

$$H^3 = H^2H = HH = H^2 = H$$

Basically, for k ¿ 1

$$H^K = H^2 H^{k-2}$$
$$H^K = H H^{k-2}$$
$$H^K = H^2 H^{k-3}$$

and so on till

$$H^K = H^2 H^{k-k}$$
$$H^K = H^2$$
$$H^K = H$$

1.3 C. To show that $(I-H)^K = I-H$ for any positive k where I is identity matrix

Let us find

$$(I-H)^2 = (I^2 + H^2 - IH - HI)$$

We know that $IA = AI = A$ and $I^2 = I$ where I is identity matrix Therefore we get

$$(I-H)^2 = I + H^2 - 2H$$

1.b we showed that $H^k = H$ for $k > 0$ Therefore

$$(I-H)^2 = I + H - 2H$$
$$(I-H)^2 = I - H$$

We get

$$(I-H)^3 = (I-H)^2(I-H) = (I-H)(I-H) = (I-H)^2 = (I-H)$$

Hence we get

$$(I-H)^k = (I-H)^2(I-H)^{k-2}$$
$$(I-H)^k = (I-H)(I-H)^{k-2}$$
$$(I-H)^k = (I-H)^2(I-H)^{k-3}$$
$$(I-H)^k = (I-H)(I-H)^{k-3}$$

and so on till

$$(I-H)^k = (I-H)(I-H)^{k-k}$$

Thus we get

$$(I-H)^k = I - H$$

1.4 D. To show that $trace(H) = d+1$

$$trace(H) = trace(X(X^TX)^{-1}X^T)$$

We know that $trace(AB) = trace(BA)$

$$trace(H) = trace(XX^T(X^TX)^{-1})$$
$$trace(H) = trace((XX^T)(X^TX)^{-1})$$
$$trace(H) = trace(X^TX(X^TX)^{-1})$$

cancelling out (X^TX) and $(X^TX)^{-1}$. Also X has dimensions $N \times (d+1)$. Therefore, X^TX with be multiplying matrix of dimension $(d+1) \times N$ with matrix $Ntimes(d+1)$ resulting in the matrix with dimensions $(d+1)$ we get

$$trace(H) = trace(I_{d+1})$$

Hence proved

2 Question 2. Prove that $\mathbb{E}[E_{in}(\mathbf{w}_{in})] = \sigma^2(1 - \frac{d+1}{N})$

2.1 A. Show that $\hat{y} = Xw^* + H\epsilon$

$$\hat{y} = Xw_{lin}$$

We know that $w_{lin} = (X^TX)^{-1}X^Ty$ Therefore

$$\hat{y} = X(X^TX)^{-1}X^Ty$$

$$\hat{y} = Hy$$

where $H = X(X^TX)^{-1}X^T$
also,

$$y = Xw^* + \epsilon$$

Therefore

$$\hat{y} = H(Xw^* + \epsilon)$$

$$\hat{y} = HXw^* + H\epsilon$$

Now

$$H = X(X^TX)^{-1}X^T$$

We get

$$\hat{y} = X(X^TX)^{-1}X^TXw^* + H\epsilon$$

$$\hat{y} = X(X^TX)^{-1}(X^TX)w^* + H\epsilon$$

Cancelling out $(X^TX)^{-1}$ and (X^TX) We get

$$\hat{y} = Xw^* + H\epsilon$$

2.2 B. Prove that $\hat{y} - y = (MATRIX)\epsilon$ and find MATRIX

$$\hat{y} - y = Xw^* + H\epsilon - (Xw^* + \epsilon)$$

$$\hat{y} - y = (H\epsilon - \epsilon)$$

$$\hat{y} - y = (H - I)\epsilon$$

2.3 C. Express E_{in}

$$E_{in} = \frac{1}{N}||Xw_{lin} - y||^2$$

Now $\hat{y} = Xw_{lin}$, we also switched the position of Xw_{lin} and y to match the answer

$$E_{in} = \frac{1}{N}||y - \hat{y}||^2$$

.... equation 1
from part B we have

$$\hat{y} - y = (I - H)\epsilon$$

we get

$$||\hat{y} - y||^2 = ||(I - H)\epsilon||^2$$

or

$$||y - \hat{y}||^2 = ||(I - H)\epsilon||^2$$

putting this in equation 1 we get

$$E_{in} = \frac{1}{N}||(I - H)\epsilon||^2$$

Now $||A||^2 = A^TA$ Thus we get

$$E_{in} = \frac{1}{N}((I-H)\epsilon)^T((I-H)\epsilon)$$

$$E_{in} = \frac{1}{N}\epsilon^T(I-H)^T(I-H)\epsilon$$

$$E_{in} = \frac{1}{N}\epsilon^T(H^T-I^T)(I-H)\epsilon$$

Now we have proved in Q1 A. that H is symmetrical i.e. $H^T = H$ and we also know that $I^T = I$
Thus we get

$$E_{in} = \frac{1}{N}\epsilon^T(I-H)(I-H)\epsilon$$

$$E_{in} = \frac{1}{N}\epsilon^T(I-H)^2\epsilon$$

We have proved that $(I-H)^k = I-H$ for any $k > 0$, Therefore we get

$$E_{in} = \frac{1}{N}\epsilon^T(I-H)\epsilon$$

2.4 D. Proving $\mathbb{E}[E_{in}(\mathbf{w}_{in})] = \sigma^2(1-\frac{d+1}{N})$ using C part

We have

$$E_{in} = \frac{1}{N}\epsilon^T(I-H)\epsilon$$

To prove

$$E[E_{in}(\mathbf{w}_{in})] = \sigma^2(1-\frac{d+1}{N})$$

Proof:

$$\mathbb{E}_D[E_{in}] = \mathbb{E}_D[\frac{1}{N}\epsilon^T(I-H)\epsilon]$$

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}\,\mathbb{E}_D[\epsilon^T I\epsilon - \epsilon^T H\epsilon]$$

Using property of expectation

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(\mathbb{E}_D[\epsilon^T I\epsilon] - \mathbb{E}_D[\epsilon^T H\epsilon])$$

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(\mathbb{E}_D[\epsilon^T \epsilon] - \mathbb{E}_D[\epsilon^T H\epsilon])$$

We will do the matrix multiplication using summation now We will write $\epsilon^T\epsilon = \sum_{i=1}^{i=N}\epsilon_i\epsilon_i$ because it is the product of The transpose of a column vector and that column vector.
We will write $\epsilon^T H\epsilon = \sum_{i=1}^{N}\sum_{j=1}^{N}\epsilon_i h_{ij}\epsilon_j$ since it is the product of 3 matrices , first is a row matrix ϵ^T , the second one is a square matrix H the third is a column matrix ϵ
We get

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(\mathbb{E}_D[\sum_{i=1}^{i=N}\epsilon_i\epsilon_i] - \mathbb{E}_D[\sum_{i=1}^{N}\sum_{j=1}^{N}\epsilon_i h_{ij}\epsilon_j])$$

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(\mathbb{E}_D[\sum_{i=1}^{i=N}\epsilon_i^2] - \mathbb{E}_D[\sum_{i=1}^{N}\sum_{j=1}^{N}\epsilon_i h_{ij}\epsilon_j])$$

Now $\mathbb{E}_D[\sum_{i=1}^{i=N}\epsilon_i^2] = \mathbb{E}_D[\epsilon_1^2+\epsilon_2^2+\epsilon_3^2+...+\epsilon_n^2] = \mathbb{E}_D[\epsilon_1^2]+\mathbb{E}_D[\epsilon_2^2]+\mathbb{E}_D[\epsilon_3^2]+...+\mathbb{E}_D[\epsilon_n^2] = \sigma^2+\sigma^2+ ...Ntimes...+\sigma^2 = N\sigma^2$

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(N\sigma^2 - \mathbb{E}_D[\sum_{i=1}^{N}\sum_{j=1}^{N}\epsilon_i h_{ij}\epsilon_j])$$

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(N\sigma^2 - \sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}_D[\epsilon_i h_{ij}\epsilon_j])$$

Now ϵ_i and ϵ_j where $i \neq j$ are independent of each other, therefore we have $E[\epsilon_i\epsilon_j] = E[\epsilon_i]E[\epsilon_j]$. Now we know that mean of ϵ is 0. Therefore $E[\epsilon_i] = E[\epsilon_j] = 0$. Now we have kept the terms having the product of ϵ_i and $\epsilon_j = 0$

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(N\sigma^2 - \sum_{i=1}^{N}\mathbb{E}_D[\epsilon_i\epsilon_i h_{ii}])$$

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(N\sigma^2 - \sum_{i=1}^{N}\mathbb{E}_D[\epsilon_i^2 h_{ii}])$$

Now since H matrix is independent of the data-set therefore we can take it out of the expectation We get,

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(N\sigma^2 - \sum_{i=1}^{N}h_{ii}\,\mathbb{E}_D[\epsilon_i^2])$$

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(N\sigma^2 - \sum_{i=1}^{N}h_{ii}\sigma^2)$$

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(N\sigma^2 - \sigma^2\sum_{i=1}^{N}h_{ii})$$

Now we know that $\sum_{i=1}^{N} h_{ii} = trace(H)$, Therefore

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}(N\sigma^2 - \sigma^2 trace(H))$$

We know that $trace(H) = d+1$ from Question 1. D Therefore we get,

$$\mathbb{E}_D[E_{in}] = \frac{1}{N}N\sigma^2 - \sigma^2(d+1))$$

Taking $\frac{1}{N}$ inside we get

$$\mathbb{E}_D[E_{in}] = \frac{(N\sigma^2 - \sigma^2(d+1)}{N}$$

$$\mathbb{E}_D[E_{in}] = \sigma^2 - \frac{\sigma^2(d+1)}{N}$$

$$\mathbb{E}_D[E_{in}] = \sigma^2(1 - \frac{d+1}{N})$$

2.5 E. To show $\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \sigma^2(1+\frac{d+1}{N})$

We know that E_{test} is generated from the test data-set. We will have similar X and w^* for E_{in} and E_{test} therefore, let us define $y' = Xw^* + \epsilon'$ (we are given the noise for E_{test} as ϵ'
Now the value of E_{test} will be the difference of $\hat{y}$ and y'
We get

$$\mathbb{E}_{test} = \frac{1}{N}||y' - \hat{y}||^2$$

Replacing the value of $\hat{y}$ from Part A and the value mentioned here for y' We get

$$E_{test} = \frac{1}{N}||(Xw^* + \epsilon') - (Xw^* + H\epsilon)||^2$$

$$E_{test} = \frac{1}{N}||\epsilon' - H\epsilon||^2$$

Now,

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \mathbb{E}_{D,\epsilon'}[\frac{1}{N}||\epsilon' - H\epsilon||^2]$$

Now, we know that $||A|| = A^T A$, therefore we get ,

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \mathbb{E}_{D,\epsilon'}[\frac{1}{N}(\epsilon' - H\epsilon)^T(\epsilon' - H\epsilon)]$$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \mathbb{E}_{D,\epsilon'}[\frac{1}{N}(\epsilon'^T - \epsilon^T H^T)(\epsilon' - H\epsilon)]$$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \mathbb{E}_{D,\epsilon'}[\frac{1}{N}(\epsilon'^T\epsilon' - \epsilon^T H^T\epsilon' - \epsilon'^T H\epsilon + \epsilon^T H^T H\epsilon)]$$

Now we know that $H^T = H$, Therefore

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \mathbb{E}_{D,\epsilon'}[\frac{1}{N}(\epsilon'^T\epsilon' - \epsilon^T H\epsilon' - \epsilon'^T H\epsilon + \epsilon^T HH\epsilon)]$$

We know that $HH = H^2 = H$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \mathbb{E}_{D,\epsilon'}[\frac{1}{N}(\epsilon'^T\epsilon' - \epsilon^T H\epsilon' - \epsilon'^T H\epsilon + \epsilon^T H\epsilon])$$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\mathbb{E}_{D,\epsilon'}[\epsilon'^T\epsilon'] - \mathbb{E}_{D,\epsilon'}[\epsilon^T H\epsilon'] - \mathbb{E}_{D,\epsilon'}[\epsilon'^T H\epsilon] + \mathbb{E}_{D,\epsilon'}[\epsilon^T H\epsilon])$$

Now, we know that ϵ' and ϵ are independent, therefore, using same concept as in previous part we can make all the terms having the term $\mathbb{E}_{D,\epsilon'}[\epsilon_i h_{ij}\epsilon'_j] = 0$ and $\mathbb{E}_{D,\epsilon'}[\epsilon'_i h_{ij}\epsilon_j] = 0$ when we do the multiplication. We get ,

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\mathbb{E}_{D,\epsilon'}[\epsilon'^T\epsilon'] + \mathbb{E}_{D,\epsilon'}[\epsilon^T H\epsilon])$$

Now we will do matrix multiplication of ϵ'^T with ϵ'
$\epsilon'^T\epsilon' = \sum_{i=1}^N \epsilon'_i\epsilon'_i = \sum_{i=1}^N \epsilon_i'^2$
and, matrix multiplication of ϵ^T with ϵ
$\epsilon^T H\epsilon = \sum_{i=1}^N \sum_{j=1}^N \epsilon_i h_{ij}\epsilon_j$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\mathbb{E}_{D,\epsilon'}[\sum_{i=1}^N \epsilon_i'^2] + \mathbb{E}_{D,\epsilon'}[\sum_{i=1}^N h_{ii}\epsilon_i^2])$$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\mathbb{E}_{D,\epsilon'}[\sum_{i=1}^N \epsilon_i'^2] + \mathbb{E}_{D,\epsilon'}[\sum_{i=1}^N \sum_{j=1}^N \epsilon_i h_{ij}\epsilon_j])$$

Taking the summation outside the expectation we have

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\sum_{i=1}^N \mathbb{E}_{D,\epsilon'}[\epsilon_i'^2] + \sum_{i=1}^N \sum_{j=1}^N \mathbb{E}_{D,\epsilon'}[\epsilon_i h_{ij}\epsilon_j])$$

Since $\mathbb{E}_{D,\epsilon'}$ is independent of h_{ij} therefore,

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\sum_{i=1}^N \mathbb{E}_{D,\epsilon'}[\epsilon_i'^2] + \sum_{i=1}^N \sum_{j=1}^N h_{ij}\,\mathbb{E}_{D,\epsilon'}[\epsilon_i\epsilon_j])$$

Now, $\mathbb{E}_{D,\epsilon'}[\epsilon_i'^2] = \sigma^2$ and using the same concept as in previous part, i.e. ϵ_i and ϵ_j where $i \neq j$, are independent , therefore, $\mathbb{E}_{D,\epsilon'}[\epsilon_i\epsilon_j] = \mathbb{E}_{D,\epsilon'}[\epsilon_i]\,\mathbb{E}_{D,\epsilon'}[\epsilon_j] = 0$ where $i \neq j$ Therefore we have

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\sum_{i=1}^N \sigma^2 + \sum_{i=1}^N h_{ii}\,\mathbb{E}_{D,\epsilon'}[\epsilon_i\epsilon_i])$$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\sum_{i=1}^{N}\sigma^2 + \sum_{i=1}^{N} h_{ii}\,\mathbb{E}_{D,\epsilon'}[\epsilon_i^2])$$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\sum_{i=1}^{N}\sigma^2 + \sum_{i=1}^{N} h_{ii}\,\mathbb{E}_{D,\epsilon'}[\epsilon_i^2])$$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}(\sum_{i=1}^{N}\sigma^2 + \sum_{i=1}^{N} h_{ii}\sigma^2)$$

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}\sigma^2(N + \sum_{i=1}^{N} h_{ii})$$

Now $\sum_{i=1}^{N} h_{ii}^2 = trace(H) = d+1$ therefore,

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \frac{1}{N}\sigma^2(N + (d+1)$$

Taking $\frac{1}{N}$ inside we get

$$\mathbb{E}_{D,\epsilon'}[E_{test}(w_{lin})] = \sigma^2(1 + \frac{d+1}{N})$$

Hence proved

3 Question 3. To prove $E_{out}(w_{lin}) = \sigma^2(1 + \frac{d+1}{N} + o(\frac{1}{N}))$

3.1 A. For a test point x, show that the error $y - g(x_n) = \epsilon - x_n^T(X^TX)^{-1}X^T\epsilon$

We used $g(x_n)$ because we need to find for a particular test point
We know that $g(x_n) = x_n^T w_{lin}$
Now, $w_{lin} = (X^TX)^{-1}X^Ty$
using this value we get

$$g(x_n) = x_n^T(X^TX)^{-1}X^Ty$$

Now $y = (Xw^* + \epsilon)$, thus we get

$$g(x_n) = x_n^T(X^TX)^{-1}X^T(Xw^* + \epsilon)$$

$$g(x_n) = x_n^T(X^TX)^{-1}X^TXw^* + x_n^T(X^TX)^{-1}X^T\epsilon$$

$$g(x_n) = x_n^Tw^* + x_n^T(X^TX)^{-1}X^T\epsilon$$

Therefore, putting these values we get

$$y_n - g(x_n) = (x_n^Tw^* + \epsilon_n) - (x_n^Tw^* + x_n^T(X^TX)^{-1}X^T\epsilon)$$

Thus we get

$$y_n - g(x_n) = \epsilon_n - x_n^T(X^TX)^{-1}X^T\epsilon$$

Hence proved

3.2 B. To prove $E_{out} = \sigma^2 + trace(\sum(X^TX)^{-1}X^T\epsilon\epsilon^TX^T(X^TX)^{-1})$

Writing equation 2.17 of the book as

$$E_{out} = E[(y - g(x_n))^2]$$

We have showed that

$$y_n - g(x_n) = \epsilon_n - x_n^T(X^TX)^{-1}X^T\epsilon$$

Therefore

$$E_{out} = E[(\epsilon_n - x_n^T(X^TX)^{-1}X^T\epsilon)^2]$$

$$E_{out} = E[(\epsilon_n - x_n^T(X^TX)^{-1}X^T\epsilon)(\epsilon_n - x_n^T(X^TX)^{-1}X^T\epsilon)^T]$$
$$E_{out} = E[(\epsilon_n - x_n^T(X^TX)^{-1}X^T\epsilon)(\epsilon_n - (x_n^T(X^TX)^{-1}X^T\epsilon)^T)]$$
$$E_{out} = E[\epsilon_n^2 - \epsilon_n(x_n^T(X^TX)^{-1}X^T\epsilon)^T - x_n^T(X^TX)^{-1}X^T\epsilon\epsilon_n + (x_n^T(X^TX)^{-1}X^T\epsilon)(x_n^T(X^TX)^{-1}X^T\epsilon)^T]$$

Since $x_n^T(X^TX)^{-1}X^T\epsilon$ is scalar, we can write $\epsilon_n(x_n^T(X^TX)^{-1}X^T\epsilon)^T = x_n^T(X^TX)^{-1}X^T\epsilon\epsilon_n$ and we will add the 2nd and the third term

$$E_{out} = E[\epsilon_n^2 - 2\epsilon_n x_n^T(X^TX)^{-1}X^T\epsilon + (x_n^T(X^TX)^{-1}X^T\epsilon)(x_n^T(X^TX)^{-1}X^T\epsilon)^T]$$

Taking E inside we get ,

$$E_{out} = E[\epsilon_n^2] - E[2\epsilon_n x_n^T(X^TX)^{-1}X^T\epsilon] + E[(x_n^T(X^TX)^{-1}X^T\epsilon)(x_n^T(X^TX)^{-1}X^T\epsilon)^T]$$

Now $E[\epsilon_n^2] = \sigma^2$ also ϵ_t is independent of x_t Thus we get

$$E_{out} = \sigma^2 - 2E[\epsilon_n]E[x_n^T(X^TX)^{-1}X^T\epsilon] + E[(x_n^T(X^TX)^{-1}X^T\epsilon)(x_n^T(X^TX)^{-1}X^T\epsilon)^T]$$

since $E[\epsilon_n] = 0$ Therefore the 2nd term in the above expression is equal to 0, we get

$$E_{out} = \sigma^2 + E[(x_n^T(X^TX)^{-1}X^T\epsilon)(x_n^T(X^TX)^{-1}X^T\epsilon)^T]$$

Now the term $(x_n^T(X^TX)^{-1}X^T\epsilon)(x_n^T(X^TX)^{-1}X^T\epsilon)^T$ is a scalar therefore we can put trace on this equation because trace(scalar) = scalar

$$E_{out} = \sigma^2 + E[trace((x_n^T(X^TX)^{-1}X^T\epsilon)(x_n^T(X^TX)^{-1}X^T\epsilon)^T)]$$

$$E_{out} = \sigma^2 + E[trace((x_n^T(X^TX)^{-1}X^T\epsilon)(\epsilon^TX(X^TX)^{-1}x_n))]$$

Now we know that trace(ABC) = trace(BCA) = trace(CAB) Therefore

$$E_{out} = \sigma^2 + E[trace(x_nx_n^T(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1})]$$
$$E_{out} = \sigma^2 + E[trace(x_nx_n^T(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1})]$$

Taking trace outside the expectation since it is a linear function

$$E_{out} = \sigma^2 + trace(E[x_nx_n^T(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}])$$

Now we know that x_n is independent of ϵ Therefore, we can separate the terms, x_n and x_n^T in the above expression

$$E_{out} = \sigma^2 + trace(E[x_nx_n^T]E[(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}])$$

We are given that $\sum = E_x[x_nx_n^T]$, we put the value in the above equation, also the term $(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}$ is a constant therefore we can remove the E

$$E_{out} = \sigma^2 + trace(\sum(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1})$$

Hence proved

3.3 C. Find $\mathbb{E}_\epsilon[\epsilon\epsilon^T]$

ϵ is a column vector of $N \times 1$ dimension , ϵ^T will be a row vector of $1 \times N$ dimension Therefore there product i.e. $\epsilon\epsilon^T$ will be of $N \times N$ dimension Now each entry of the matrix of the form $\epsilon_i\epsilon_j$ where $i \neq j$ will be 0 because ϵ_i will be independent of ϵ_j where $i \neq j$ Hence it will just have terms of form ϵ_i^2 in the diagonal and rest of the terms will be 0
We can push the expectation inside the matrix such that we find the expectation of each element of the matrix. Only the diagonal elements will have a non zero expectation $= \sigma^2$ because ($= E[\epsilon_i^2]$) and rest of the elements will be 0 . Therefore the resultant matrix will be of form $\sigma^2 I_N$ where I_N is an identity matrix of $N \times N$ size

$$\mathbb{E}_\epsilon[\epsilon\epsilon^T] = \sigma^2 I_N$$

3.4 D. $E_{out} = \sigma^2 + \frac{\sigma^2}{N} trace(\sum(\frac{1}{N}X^TX)^{-1})$

From B we have

$$E_{out} = \sigma^2 + trace(\sum(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1})$$

Now we will take expectation over ϵ ,

$$E_{out} = E_\epsilon[\sigma^2 + trace(\sum(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1})]$$

$$E_{out} = E_\epsilon[\sigma^2] + E_\epsilon[trace(\sum(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1})]$$

$$E_{out} = \sigma^2 + E_\epsilon[trace(\sum(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1})]$$

$\sum(X^TX)^{-1}X^T$ and $X(X^TX)^{-1})$ are constant quantities therefore we can push expectation inside

$$E_{out} = \sigma^2 + trace(E[\sum(X^TX)^{-1}X^T]E[\epsilon\epsilon^T]E[X(X^TX)^{-1})]$$

since $\sum(X^TX)^{-1}X^T$ and $X(X^TX)^{-1})$ are constant quantities,

$$E_{out} = \sigma^2 + trace(\sum(X^TX)^{-1}X^TE[\epsilon\epsilon^T]X(X^TX)^{-1})$$

From C part

$$E_{out} = \sigma^2 + trace(\sum(X^TX)^{-1}X^T\sigma^2I_NX(X^TX)^{-1})$$

$$E_{out} = \sigma^2 + \sigma^2 trace(\sum(X^TX)^{-1}X^TX(X^TX)^{-1})$$

$$E_{out} = \sigma^2 + \sigma^2 trace(\sum(X^TX)^{-1})$$

Now to match the answer we will divide by N outside and divide by N^{-1} inside

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N} trace(\sum(\frac{1}{N}X^TX)^{-1})$$

We are given that $\frac{1}{N}X^TX = \sum$ Putting its value,

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N} trace(\sum(\sum)^{-1})$$

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N} trace(I_{d+1})$$

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N}(d+1)$$

$$E_{out} = \sigma^2(1 + \frac{d+1}{N})$$

3.5 E. $E_{out} = \sigma^2(1 + \frac{d+1}{N} + o(\frac{1}{N}))$

According to law of large numbers $\frac{1}{N}X^TX$ converges in probability to $\sum$ and $(\frac{1}{N}X^TX)^{-1}$ converges to $(\sum)^{-1}$ Now the multiplication of $\frac{1}{N}X^TX$ and $(\frac{1}{N}X^TX)^{-1}$ will converge to $\sum(\sum)^{-1}$

Thus we can write $\frac{1}{N}X^TX \times (\frac{1}{N}X^TX)^{-1} = \sum(\sum)^{-1} + o(1) = I + o(1)$ Therefore we get ,

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N} trace(\sum(\frac{1}{N}X^TX)^{-1})$$

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N} trace(I + o(1))$$

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N}(trace(I) + o(1)))$$

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N}((d+1) + o(1)))$$

$$E_{out} = \sigma^2(1 + \frac{d+1}{N} + \frac{o(1)}{N})$$