Amaiya Singhal 2021CS50598

COL341 Homework 1

- 1 a

Proof.

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

Taking transpose of H

$$H^T = (X(X^TX)^{-1}X^T)^T$$
$$H^T = (X^T)^T(X(X^TX)^{-1})^T \qquad (AB)^T = B^TA^T$$
$$H^T = X((X^TX)^{-1})^TX^T \qquad (A^T)^T = A$$
$$H^T = X((X^TX)^T)^{-1}X^T \qquad (A^{-1})^T = (A^T)^{-1}$$
$$H^T = X((X^T)(X^T)^T)^{-1}X^T \qquad (AB)^T = B^TA^T$$
$$H^T = X((X^T)X)^{-1}X^T \qquad (X^T)^T = X$$
$$H^T = X(X^TX)^{-1}X^T = H$$

As H^T is equal to H, we can conclude that H is symmetric. □

- 1 b

Proof. *(by Induction on exponent of the matrix)*
Inductive Hypothesis: $H^K = H$ for some positive integer K
Base Case : K = 1, $H^1 = H$. Hence base case holds
Inductive Step : To show $H^{K+1} = H$

$$H^{K+1} = H.H^K$$
$$H^{K+1} = H.H \qquad \textit{From the Inductive Hypothesis } H^K = H$$
$$H^{K+1} = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T \qquad \textit{Substituting Value of } H$$
$$H^{K+1} = X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T \qquad \textit{Associativity}$$
$$H^{K+1} = X((X^TX)^{-1}(X^TX))(X^TX)^{-1}X^T \qquad \textit{Associativity}$$
$$H^{K+1} = X(I)(X^TX)^{-1}X^T \qquad A.A^{-1} = I$$
$$H^{K+1} = X(X^TX)^{-1}X^T = H \qquad A.I = A$$
$$H^{K+1} = H$$

□

- 1 c

Proof. (by Induction on exponent of the matrix)
Inductive Hypothesis: $(I-H)^K = I - H$ *for some positive integer* K
Base Case : $K = 1$, $(I-H)^1 = I - H$. *Hence base case holds*
Inductive Step : *To show* $(I-H)^{K+1} = I - H$

$$(I-H)^{K+1} = (I-H).(I-H)^K$$
$$(I-H)^{K+1} = (I-H).(I-H) \qquad From\ the\ I.H.\ (I-H)^K = I - H$$
$$(I-H)^{K+1} = I.(I-H) - H.(I-H) \qquad Distributivity$$
$$(I-H)^{K+1} = I.I - I.H - H.I + H.H \qquad Distributivity$$
$$(I-H)^{K+1} = I - H - H + H^2 \qquad A.I = A$$
$$(I-H)^{K+1} = I - H - H + H \qquad H^K = H\ for\ positive\ integer\ K\ (Q1.b)$$
$$(I-H)^{K+1} = I - H \qquad A + B - B = A$$

□

- 1 d

Proof.

$$Tr(H) = Tr(X(X^TX)^{-1}X^T)$$
$$Tr(H) = Tr((X(X^TX)^{-1})(X^T)) \qquad Associativity$$
$$Tr(H) = Tr((X)^T(X(X^TX)^{-1})) \qquad Tr(AB) = Tr(BA)$$
$$Tr(H) = Tr(X^TX(X^TX)^{-1}) \qquad Associativity$$
$$Tr(H) = Tr((X^TX)(X^TX)^{-1}) \qquad Associativity$$
$$Tr(H) = Tr(I_{d+1}) \qquad A.A^{-1} = I$$
$$Tr(H) = d + 1 \qquad Tr(I_k) = k$$

□

- 2 a

Proof.

$$\mathbf{y} = X\mathbf{w}^* + \epsilon$$

$$\begin{aligned}
\hat{\mathbf{y}} &= X\mathbf{w}_{lin} \\
\hat{\mathbf{y}} &= X(X^TX)^{-1}X^Ty && Substituting\ Value\ of\ \mathbf{w}_{lin} \\
\hat{\mathbf{y}} &= X(X^TX)^{-1}X^T(X\mathbf{w}^* + \epsilon) && Substituting\ Value\ of\ y \\
\hat{\mathbf{y}} &= X(X^TX)^{-1}X^TX\mathbf{w}^* + X(X^TX)^{-1}X^T\epsilon && Distributivity \\
\hat{\mathbf{y}} &= X(X^TX)^{-1}(X^TX)\mathbf{w}^* + X(X^TX)^{-1}X^T\epsilon && Associativity \\
\hat{\mathbf{y}} &= X\mathbf{w}^* + X(X^TX)^{-1}X^T\epsilon && A.A^{-1} = I \\
\hat{\mathbf{y}} &= X\mathbf{w}^* + H\epsilon && X(X^TX)^{-1}X^T = H
\end{aligned}$$

□

- 2 b

Proof.

$$\hat{\mathbf{y}} - \mathbf{y} = (X\mathbf{w}^* + H\epsilon) - (X\mathbf{w}^* + \epsilon) \qquad Substituting\ from\ 2\ a$$

$$\begin{aligned}
\hat{\mathbf{y}} - \mathbf{y} &= H\epsilon - \epsilon \\
\hat{\mathbf{y}} - \mathbf{y} &= (H - I)\epsilon && Distributivity \\
&The\ matrix\ is\ H - I
\end{aligned}$$

□

- 2 c

Solution.

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\mathbf{w}_{lin}^TX^TX\mathbf{w}_{lin} - 2\mathbf{w}_{lin}^TX^T\mathbf{y} + \mathbf{y}^T\mathbf{y})$$

$$\begin{aligned}
E_{in}(\mathbf{w}_{lin}) &= \frac{1}{N}((X\mathbf{w}_{lin})^TX\mathbf{w}_{lin} - 2(X\mathbf{w}_{lin})^T\mathbf{y} + \mathbf{y}^T\mathbf{y}) && (AB)^T = B^TA^T \\
E_{in}(\mathbf{w}_{lin}) &= \frac{1}{N}((H\mathbf{y})^TH\mathbf{y} - 2(H\mathbf{y})^T\mathbf{y} + \mathbf{y}^T\mathbf{y}) && (X\mathbf{w}_{lin}) = (H\mathbf{y}) \\
E_{in}(\mathbf{w}_{lin}) &= \frac{1}{N}(\mathbf{y}^TH^TH\mathbf{y} - 2\mathbf{y}^TH^T\mathbf{y} + \mathbf{y}^T\mathbf{y}) && (AB)^T = B^TA^T \\
E_{in}(\mathbf{w}_{lin}) &= \frac{1}{N}(\mathbf{y}^TH.H\mathbf{y} - 2\mathbf{y}^TH\mathbf{y} + \mathbf{y}^T\mathbf{y}) && H\ is\ symmetric \\
E_{in}(\mathbf{w}_{lin}) &= \frac{1}{N}(\mathbf{y}^TH^2\mathbf{y} - 2\mathbf{y}^TH\mathbf{y} + \mathbf{y}^T\mathbf{y}) && H.H = H^2 \\
E_{in}(\mathbf{w}_{lin}) &= \frac{1}{N}(\mathbf{y}^T(H^2 - 2H + I)\mathbf{y})
\end{aligned}$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\mathbf{y}^T(I-H)^2\mathbf{y}) \qquad (I-A)^2 = I^2 + A^2 - 2A$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\mathbf{y}^T(I-H)\mathbf{y}) \qquad \textit{From Q1 c}$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\mathbf{y}^T(\mathbf{y}-H\mathbf{y}))$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\mathbf{y}^T(\mathbf{y}-\hat{\mathbf{y}})) \qquad \hat{\mathbf{y}} = H\mathbf{y}$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\mathbf{y}^T(I-H)\epsilon) \qquad \textit{From Q2 b}$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}((\mathbf{y}^T-\mathbf{y}^T H)\epsilon)$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}((\mathbf{y}^T-(H\mathbf{y})^T)\epsilon) \qquad (AB)^T = B^T A^T$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}((\mathbf{y}^T-(\hat{\mathbf{y}})^T)\epsilon) \qquad \hat{\mathbf{y}} = H\mathbf{y}$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}((\mathbf{y}-\hat{\mathbf{y}})^T\epsilon) \qquad (A-B)^T = A^T - B^T$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(((I-H)\epsilon)^T\epsilon) \qquad \textit{From Q2 b}$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\epsilon^T(I-H)^T\epsilon) \qquad (AB)^T = B^T A^T$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\epsilon^T(I-H^T)\epsilon) \qquad (A-B)^T = A^T - B^T$$

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\epsilon^T(I-H)\epsilon) \qquad \textit{H is symmetric}$$

- 2 d

Proof.

$$E_{in}(\mathbf{w}_{lin}) = \frac{1}{N}(\epsilon^T(I-H)\epsilon) \qquad \textit{From Q2 c}$$

$$E_D[E_{in}(\mathbf{w}_{lin})] = E_D[\frac{1}{N}(\epsilon^T(I-H)\epsilon)]$$

$$E_D[E_{in}(\mathbf{w}_{lin})] = \frac{1}{N}E_D[(\epsilon^T(I-H)\epsilon)] \qquad E[kX] = kE[X]$$

$$E_D[E_{in}(\mathbf{w}_{lin})] = \frac{1}{N}E_D[(\epsilon^T\epsilon - \epsilon^T H\epsilon)] \qquad \textit{Distributivity}$$

$$E_D[E_{in}(\mathbf{w}_{lin})] = \frac{1}{N}(E_D[\epsilon^T\epsilon] - E_D[\epsilon^T H\epsilon]) \qquad E[X-Y] = E[X] - E[Y]$$

Calculating the expectations :

$$E_D[\epsilon^T\epsilon] = E_D[\sum_{i=1}^{N}\epsilon_i^2]$$

$$E_D[\epsilon^T\epsilon] = \sum_{i=1}^{N}E_D[\epsilon_i^2]$$

$$E_D[\epsilon^T\epsilon] = \sum_{i=1}^{N}\sigma^2$$

$$E_D[\epsilon^T\epsilon] = N\sigma^2 \qquad - 1.$$

$$E_D[\epsilon^T H\epsilon] = E_D[(\epsilon_1\epsilon_2...\epsilon_N)\begin{pmatrix} a_{11} & a_{12} & ... & a_{1N} \\ a_{21} & a_{22} & ... & a_{2N} \\ ... & ... & ... & ... \\ a_{N1} & a_{N2} & ... & a_{NN} \end{pmatrix}(\epsilon_1\epsilon_2...\epsilon_N)^T]$$

$$E_D[\epsilon^T H\epsilon] = E_D[\sum_{i=1}^{N}\epsilon_i(\sum_{j=1}^{N}\epsilon_j a_{ji})]$$

$$E_D[\epsilon^T H\epsilon] = E_D[\sum_{i=1}^{N}\sum_{j=1}^{N}\epsilon_i\epsilon_j a_{ji}] \qquad \textit{Rearranging}$$

$$E_D[\epsilon^T H\epsilon] = \sum_{i=1}^{N}\sum_{j=1}^{N}E_D[\epsilon_i\epsilon_j a_{ji}] \qquad \sum_i E[X_i] = E[\sum_i X_i]$$

$$E_D[\epsilon^T H\epsilon] = \sum_{i=1}^{N}E_D[\epsilon_i^2 a_{ii}] + \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}E_D[\epsilon_i\epsilon_j a_{ji}]$$

$$E_D[\epsilon^T H\epsilon] = \sum_{i=1}^{N} E_D[\epsilon_i^2]E_D[a_{ii}] + \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} E_D[\epsilon_i]E_D[\epsilon_j]E_D[a_{ji}] \quad (\epsilon_i, \epsilon_j \textit{ independent})$$

$$E_D[\epsilon^T H\epsilon] = \sum_{i=1}^{N} \sigma^2 E_D[a_{ii}] + \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} 0 \qquad E_D[\epsilon^2] = \sigma^2,\ E_D[\epsilon_i] = 0$$

$$E_D[\epsilon^T H\epsilon] = \sigma^2 \sum_{i=1}^{N} E_D[a_{ii}] \qquad E[kX] = kE[X]$$

$$E_D[\epsilon^T H\epsilon] = \sigma^2 \sum_{i=1}^{N} a_{ii} \qquad E_D[a_{ij}] = a_{ij}$$

$$E_D[\epsilon^T H\epsilon] = \sigma^2 Tr(H) \qquad Tr(H) = \sum_{i=1}^{N} a_{ii}$$

$$E_D[\epsilon^T H\epsilon] = \sigma^2(d+1) \qquad - 2.$$

$$E_D[E_{in}(\mathbf{w}_{lin})] = \frac{1}{N}(\sigma^2 N - \sigma^2(d+1)) \qquad \textit{Substituting 1. and 2.}$$

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

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

□

- 2 e

Proof.

$$E_{Test}(\mathbf{w}_{lin}) = \frac{1}{N}||\mathbf{y}' - \hat{\mathbf{y}}||^2$$

$$\mathbf{y}' - \hat{\mathbf{y}} = (H\mathbf{w}^* + H\epsilon) - (H\mathbf{w}^* + \epsilon')$$

$$\mathbf{y}' - \hat{\mathbf{y}} = H\epsilon - \epsilon'$$

$$E_{Test}(\mathbf{w}_{lin}) = \frac{1}{N}||H\epsilon - \epsilon'||^2$$

$$E_{Test}(\mathbf{w}_{lin}) = \frac{1}{N}(H\epsilon - \epsilon')^T(H\epsilon - \epsilon')$$

$$E_{Test}(\mathbf{w}_{lin}) = \frac{1}{N}((H\epsilon)^T - \epsilon'^T)(H\epsilon - \epsilon') \qquad (A-B)^T = A^T - B^T$$

$$E_{Test}(\mathbf{w}_{lin}) = \frac{1}{N}(\epsilon^T H^T - \epsilon'^T)(H\epsilon - \epsilon') \qquad (AB)^T = B^T A^T$$

$$E_{Test}(\mathbf{w}_{lin}) = \frac{1}{N}(\epsilon^T H - \epsilon'^T)(H\epsilon - \epsilon') \qquad H^T = H$$

$$E_{Test}(\mathbf{w}_{lin}) = \frac{1}{N}(\epsilon^T H.H\epsilon - \epsilon'^T H\epsilon - \epsilon^T H\epsilon' + \epsilon'^T\epsilon') \qquad \textit{Distributivity}$$

$$E_{Test}(\mathbf{w}_{lin}) = \frac{1}{N}(\epsilon^T H^2 \epsilon - \epsilon'^T H \epsilon - \epsilon^T H \epsilon' + \epsilon'^T \epsilon')$$

$$E_{Test}(\mathbf{w}_{lin}) = \frac{1}{N}(\epsilon^T H \epsilon - \epsilon'^T H \epsilon - \epsilon^T H \epsilon' + \epsilon'^T \epsilon') \qquad H^k = H$$

$$E_D[E_{Test}(\mathbf{w}_{lin})] = E_D[\frac{1}{N}(\epsilon^T H \epsilon - \epsilon'^T H \epsilon - \epsilon^T H \epsilon' + \epsilon'^T \epsilon')]$$

$$E_D[E_{Test}(\mathbf{w}_{lin})] = \frac{1}{N} E_D[\epsilon^T H \epsilon - \epsilon'^T H \epsilon - \epsilon^T H \epsilon' + \epsilon'^T \epsilon'] \qquad E[kX] = kE[X]$$

$$E_D[E_{Test}(\mathbf{w}_{lin})] = \frac{1}{N}(E_D[\epsilon^T H \epsilon] - E_D[\epsilon'^T H \epsilon] - E_D[\epsilon^T H \epsilon'] + E_D[\epsilon'^T \epsilon']) \quad Linearity$$

$From\ Q2\ d,\ we\ have$

$$E_D[\epsilon^T H \epsilon] = \sigma^2(d+1) \qquad -1.$$

$$E_D[\epsilon'^T \epsilon'] = N\sigma^2 \qquad -2.$$

$$E_D[\epsilon^T H \epsilon'] = \sum_{i=1}^{N} E_D[\epsilon_i] E_D[\epsilon'_i] E_D[a_{ii}] + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} E_D[\epsilon'_i] E_D[\epsilon_j] E_D[a_{ji}]$$

$$E_D[\epsilon^T H \epsilon'] = \sum_{i=1}^{N} 0 + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} 0$$

$$E_D[\epsilon^T H \epsilon'] = 0 \qquad -3.$$

$$Similarly\ E_D[\epsilon'^T H \epsilon] = 0 \qquad -4.$$

$$E_D[E_{Test}(\mathbf{w}_{lin})] = \frac{1}{N}(\sigma^2(d+1) + 0 + 0 + \sigma^2 N) \qquad Substituting\ 1,2,3\ and\ 4$$

$$E_D[E_{Test}(\mathbf{w}_{lin})] = \frac{1}{N}\sigma^2(N + (d+1))$$

$$E_D[E_{Test}(\mathbf{w}_{lin})] = \sigma^2(1 + \frac{d+1}{N})$$

$\square$

- 3 a

Proof.

$$\text{The train dataset is } \mathbf{y} = X\mathbf{w}^* + \boldsymbol{\epsilon}$$

$\mathbf{w}_{lin} = (X^TX)^{-1}X^T\mathbf{y}$

The prediction for the test data $\mathbf{x}$ *is* $g(\mathbf{x})$

$g(\mathbf{x}) = \mathbf{x}^T\mathbf{w}_{lin}$

$g(\mathbf{x}) = \mathbf{x}^T((X^TX)^{-1}X^T\mathbf{y})$

$g(\mathbf{x}) = \mathbf{x}^T((X^TX)^{-1}X^T(X\mathbf{w}^* + \boldsymbol{\epsilon}))$

$g(\mathbf{x}) = \mathbf{x}^T((X^TX)^{-1}X^TX\mathbf{w}^* + (X^TX)^{-1}X^T\boldsymbol{\epsilon})$ *Distributivity*

$g(\mathbf{x}) = \mathbf{x}^T((X^TX)^{-1}(X^TX)\mathbf{w}^* + (X^TX)^{-1}X^T\boldsymbol{\epsilon})$

$g(\mathbf{x}) = \mathbf{x}^T(\mathbf{w}^* + (X^TX)^{-1}X^T\boldsymbol{\epsilon})$ $AA^{-1} = I$

$g(\mathbf{x}) = \mathbf{x}^T\mathbf{w}^* + \mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}$ *Distributivity*

$\mathbf{y}_{test} = \mathbf{x}^T\mathbf{w}^* + \epsilon$

$Error \;=\; \mathbf{y}_{test} - g(\mathbf{x})$

$\mathbf{y}_{test} - g(\mathbf{x}) = \mathbf{x}^T\mathbf{w}^* + \epsilon - (\mathbf{x}^T\mathbf{w}^* + \mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})$

$\mathbf{y}_{test} - g(\mathbf{x}) = \epsilon - \mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}$

□

- 3 b

Proof.

$E_{out} = E_{\mathbf{x},\epsilon}[(\mathbf{y}_{test} - g(\mathbf{x}))^2] = E_{\mathbf{x},\epsilon}[(\epsilon - \mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})^2]$

$E_{out} = E_{\mathbf{x},\epsilon}[(\epsilon - \mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})^2]$

$E_{out} = E_{\mathbf{x},\epsilon}[\epsilon^2 - 2\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon} + (\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})(\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})^T]$

$E_{out} = E_{\mathbf{x},\epsilon}[\epsilon^2 - 2\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon} + (\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})(\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x})]$

$E_{out} = E_{\mathbf{x},\epsilon}[\epsilon^2 - 2\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon} + \mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x}]$

$E_{out} = E_{\mathbf{x},\epsilon}[\epsilon^2] - E_{\mathbf{x},\epsilon}[2\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}] + E_{\mathbf{x},\epsilon}[\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x}]$

Using $E_{\mathbf{x},\epsilon}[\epsilon^2] = \sigma^2,\; E_{\mathbf{x},\epsilon}[\epsilon] = 0$

$E_{out} = \sigma^2 - E_{\mathbf{x},\epsilon}[2\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}] + E_{\mathbf{x},\epsilon}[\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x}]$

$E_{out} = \sigma^2 - 0 + E_{\mathbf{x},\epsilon}[\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x}]$

Since $\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x}$ *is a scalar*

$E_{out} = \sigma^2 + E_{\mathbf{x},\epsilon}[Tr(\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x})]$

$E_{out} = \sigma^2 + E_{\mathbf{x},\epsilon}[Tr(\mathbf{x}\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1})]$ $Tr(AB) = Tr(BA)$

$$E_{out} = \sigma^2 + Tr(E_{\mathbf{x},\epsilon}[\mathbf{x}\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1}]) \qquad Tr(E[X]) = E[Tr(X)]$$

$$E_{out} = \sigma^2 + Tr(E_{\mathbf{x},\epsilon}[\mathbf{x}\mathbf{x}^T]E_{\mathbf{x},\epsilon}[(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1}]) \qquad Independent$$

$$E_{out} = \sigma^2 + Tr(\Sigma E_{\mathbf{x},\epsilon}[(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1}]) \qquad \Sigma = E_{\mathbf{x},\epsilon}[\mathbf{x}\mathbf{x}^T]$$

$$E_{out} = \sigma^2 + Tr(\Sigma(X^TX)^{-1}X^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^TX(X^TX)^{-1})$$

□

- 3 c

Solution.

$$E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = E_{\boldsymbol{\epsilon}}[\begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ ... \\ \boldsymbol{\epsilon}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & ... & \boldsymbol{\epsilon}_N \end{pmatrix}]$$

$$E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = E_{\boldsymbol{\epsilon}}[\begin{pmatrix} \boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_2 & ... & \boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_N \\ \boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_2 & ... & \boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_N \\ ... & ... & ... & ... \\ \boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_2 & ... & \boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_N \end{pmatrix}]$$

$$Expectation\ of\ a\ matrix\ is\ the\ matrix\ of\ the\ expectation\ of\ its\ elements$$

$$E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \begin{pmatrix} E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1] & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_2] & ... & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_N] \\ E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_1] & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_2] & ... & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_2\boldsymbol{\epsilon}_N] \\ ... & ... & ... & ... \\ E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_1] & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_2] & ... & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_N] \end{pmatrix}$$

$$E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \begin{pmatrix} E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_1^2] & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_1]E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_2] & ... & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_1]E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_N] \\ E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_2]E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_1] & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_2^2] & ... & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_2]E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_N] \\ ... & ... & ... & ... \\ E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_N]E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_1] & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_N]E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_2] & ... & E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_N^2] \end{pmatrix} (\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j\ are\ independent)$$

$$E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \begin{pmatrix} \sigma^2 & 0 & 0 & ... & 0 \\ 0 & \sigma^2 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & ... & 0 & \sigma^2 & 0 \\ 0 & 0 & ... & 0 & \sigma^2 \end{pmatrix} \qquad E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_i] = 0,\ E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}_i^2] = \sigma^2$$

$$E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & ... & 0 & 1 & 0 \\ 0 & 0 & ... & 0 & 1 \end{pmatrix}$$

$$E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \sigma^2 I$$

- 3 d

Proof.

$$E_{out} = E_{\boldsymbol{\epsilon}}[(\mathbf{y}_{test} - g(\mathbf{x}))^2] = E_{\boldsymbol{\epsilon}}[(\epsilon - \mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})^2]$$

$$E_{out} = E_{\boldsymbol{\epsilon}}[(\epsilon - \mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})^2]$$

$$E_{out} = E_{\boldsymbol{\epsilon}}[\epsilon^2 - 2\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon} + (\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})^T(\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})]$$

$$E_{out} = E_{\boldsymbol{\epsilon}}[\epsilon^2 - 2\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon} + (\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x})(\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})]$$

$$E_{out} = E_{\boldsymbol{\epsilon}}[\epsilon^2] - E_{\boldsymbol{\epsilon}}[2\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}] + E_{\boldsymbol{\epsilon}}[(\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x})(\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon})]$$

$$E_{out} = E_{\boldsymbol{\epsilon}}[\epsilon^2] - 0 + E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^TX(X^TX)^{-1}\mathbf{x}\mathbf{x}^T(X^TX)^{-1}X^T\boldsymbol{\epsilon}]$$

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

$$E_{out} = \sigma^2 + E_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^TA\boldsymbol{\epsilon}] \qquad \textit{let } A = X(X^TX)^{-1}\mathbf{x}\mathbf{x}^T(X^TX)^{-1}X^T$$

$$E_{out} = \sigma^2 + \sigma^2E_{\boldsymbol{\epsilon}}[Tr(A)] \qquad \textit{similar to Q2 d}$$

$$E_{out} = \sigma^2 + \sigma^2E_{\boldsymbol{\epsilon}}[Tr(X(X^TX)^{-1}\mathbf{x}\mathbf{x}^T(X^TX)^{-1}X^T)]$$

$$E_{out} = \sigma^2 + \sigma^2E_{\boldsymbol{\epsilon}}[Tr(X^TX(X^TX)^{-1}\mathbf{x}\mathbf{x}^T(X^TX)^{-1})] \qquad Tr(AB) = Tr(BA)$$

$$E_{out} = \sigma^2 + \sigma^2E_{\boldsymbol{\epsilon}}[Tr((X^TX)(X^TX)^{-1}\mathbf{x}\mathbf{x}^T(X^TX)^{-1})] \qquad \textit{Associativity}$$

$$E_{out} = \sigma^2 + \sigma^2E_{\boldsymbol{\epsilon}}[Tr(\mathbf{x}\mathbf{x}^T(X^TX)^{-1})] \qquad AA^{-1} = I$$

$$E_{out} = \sigma^2 + \sigma^2Tr(E_{\boldsymbol{\epsilon}}[\mathbf{x}\mathbf{x}^T(X^TX)^{-1}]) \qquad E[Tr(A)] = Tr(E[X])$$

$$E_{out} = \sigma^2 + \sigma^2Tr(E_{\boldsymbol{\epsilon}}[\mathbf{x}\mathbf{x}^T]E_{\boldsymbol{\epsilon}}[(X^TX)^{-1}]) \qquad \textit{Independent}$$

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

$$E_{out} = \sigma^2 + \sigma^2Tr(E_{\boldsymbol{\epsilon}}[\mathbf{x}\mathbf{x}^T]E_{\boldsymbol{\epsilon}}[(X^TX)^{-1}]) \qquad \Sigma = E_{\mathbf{x},\epsilon}[\mathbf{x}\mathbf{x}^T]$$

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

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N}Tr(\Sigma N(X^TX)^{-1}) \qquad Tr(NX) = NTr(X)$$

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N}Tr(\Sigma(\frac{1}{N}X^TX)^{-1}) \qquad (\frac{A}{\lambda})^{-1} = \lambda(A)^{-1}$$

Proved.

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N}Tr(\Sigma(\Sigma)^{-1}) \qquad If\ (\frac{1}{N}X^TX) = \Sigma$$

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

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

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

□

- 3 e

Proof.

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N} Tr(\Sigma(\frac{1}{N} X^T X)^{-1})$$

$$(\frac{1}{N} X^T X)^{-1} \xrightarrow{\text{P}} \Sigma^{-1} \qquad \textit{Convergence in Probability}$$

$$P(|(\frac{1}{N} X^T X)^{-1} - \Sigma^{-1}| < \delta) = 1$$

$$(\frac{1}{N} X^T X)^{-1} = \Sigma^{-1} + o(1)$$

$$\Sigma(\frac{1}{N} X^T X)^{-1} = \Sigma\Sigma^{-1} + \Sigma o(1)$$

$$\Sigma(\frac{1}{N} X^T X)^{-1} = I + \Sigma o(1)$$

$$E_{out} = \sigma^2 + \frac{\sigma^2}{N} Tr(I + \Sigma o(1)) \qquad AA^{-1} = I$$

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

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

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

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

□