Linear Regression

Professor Jie Peng, PhD

Department of Statistics

University of California, Davis

Geometric Interpretation

Column Space of the Design Matrix

$$\begin{array}{cc} & \begin{array}{cc} 1x & \bar{x} \end{array} \\ \underset{n\times 2}{X} = & \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \end{array}$$

- $\mathbf{1}_n$: $n \times 1$ vector of ones; $\mathbf{x} = (X_1, \cdots, X_n)^T$: $n \times 1$ vector of X values.
- Design matrix: $\mathbf{X} = (\mathbf{1}_n, \mathbf{x})$.
- $\mathrm{col}(\mathrm{X})$: linear subspace of $\mathbb{R}^n$ generated by the columns of $\mathbf{X}$

$$\mathrm{col}(\mathrm{X}) = \{\mathbf{v} \in \mathbb{R}^{\mathrm{n}} : \exists \mathrm{c}_0, \mathrm{c}_1 \in \mathrm{R}, \text{s.t.}, \mathbf{v} = \mathrm{c}_0 \mathbf{1}_{\mathrm{n}} + \mathrm{c}_1 \mathbf{x}\}.$$

Projection to $\mathrm{col}(\mathrm{X})$

(orthogonally)

$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ projects a vector in $\mathbb{R}^n$ to $\mathrm{col}(\mathrm{X})$: For any $\mathbf{w} \in \mathbb{R}^n$

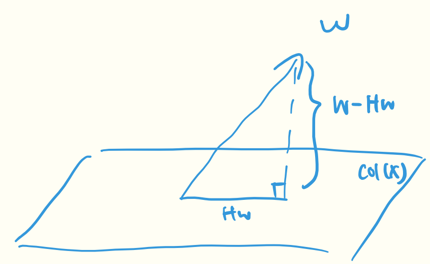
nxn belong column space of X nx1

project
- $\mathbf{Hw} \in \mathrm{col}(\mathrm{X})$, i.e., there exists $c_0, c_1 \in \mathbf{R}$ such that

  $\mathbf{Hw} = c_0\mathbf{1}_n + c_1\mathbf{x}$. can be written as linear combination of 1_n, x

orthogonal
- $\mathbf{w} - \mathbf{Hw} \perp \mathrm{col}(\mathrm{X})$, i.e., for any $\mathbf{v} \in \mathrm{col}(\mathrm{X})$, the inner product

  $< \mathbf{w} - \mathbf{Hw}, \mathbf{v} >= (\mathbf{w} - \mathbf{Hw})^T\mathbf{v} = 0$.



(i) projection

$H = X(X^TX)^{-1}X^T$, $w \in R^n$, $X = (1_n, \vec{x})$ $c = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}$

$Hw = X \cdot \underbrace{(X^TX)^{-1}X^T w}_{\vec{c}\ (2\times1)} = X \cdot \vec{c} = c_0 1_n + c_1 \vec{x}$ def. col(X)

(ii) orthogonal

for any $\vec{v} \in col(X)$, $\exists$ (exists) $\vec{c} \in R^2$,

st. $\vec{v} = X\vec{c}$ (def of col(X))

so $\langle w - Hw, \vec{v} \rangle = (w - Hw)^T \cdot v$

$= w^T (I_n - H) \cdot X\vec{c}$

$= 0$ [scalar]

$w - Hw$
$= (I_n - H) w$

$(w - Hw)^T$
$= w^T \cdot (I_n - H)^T$
$= w^T \cdot (I_n - H)$

HX
$= X(X^TX)^{-1} \cdot X^TX$
$= X$

$(I_n - H) X$
$= X - HX$
$= X - X$
$= \emptyset$ [n×2]

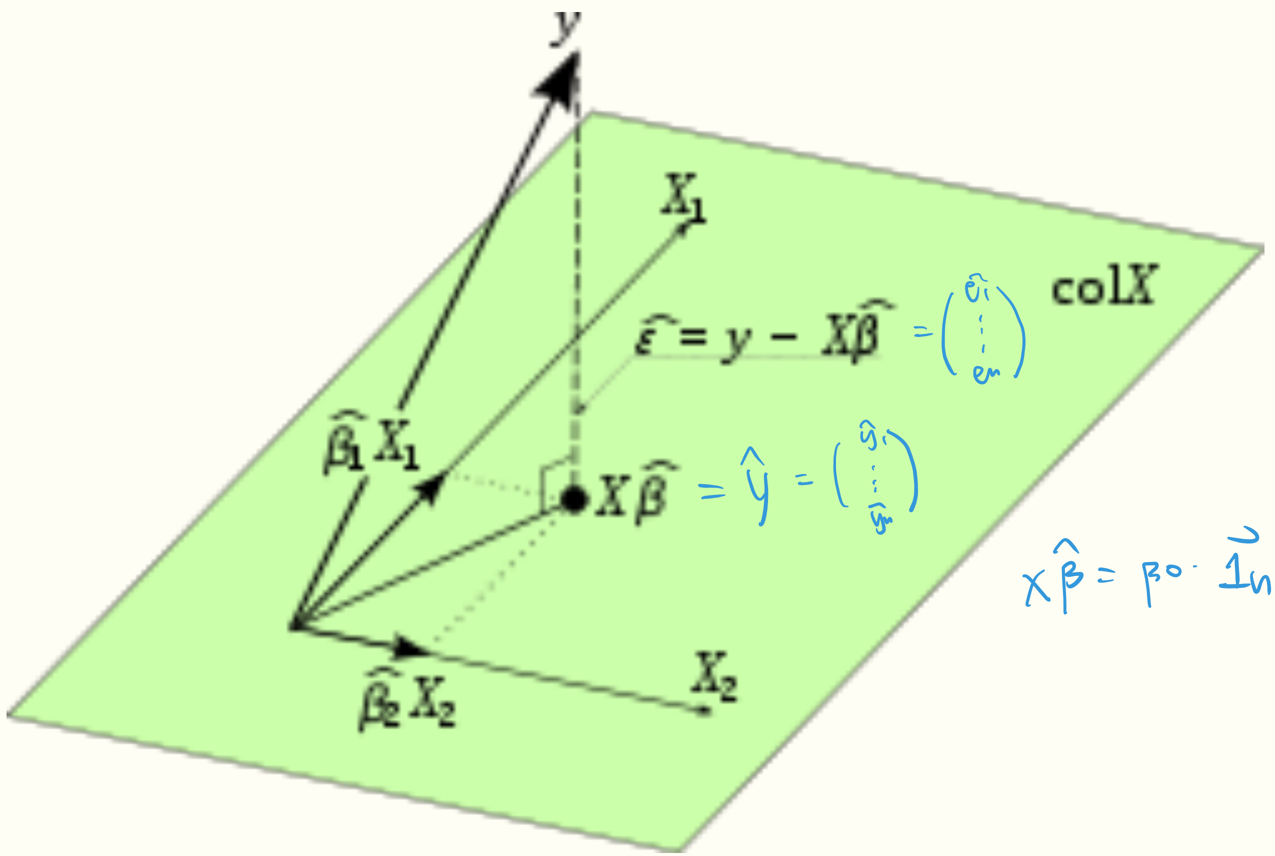
Fitted Values and Residuals

- $\widehat{\mathbf{Y}} = \mathbf{HY} = \hat{\beta}_0 \mathbf{1}_n + \hat{\beta}_1 \mathbf{x} \in \mathrm{col}(\mathrm{X})$
- $\mathbf{e} = \mathbf{Y} - \mathbf{HY} \perp \mathrm{col}(\mathrm{X})$
- Since $\mathbf{1}_n, \mathbf{x}, \widehat{\mathbf{Y}} \in \mathrm{col}(\mathrm{X})$, so

$$\left\{ \begin{aligned} < \mathbf{e}, \mathbf{1}_n > &= \sum_{i=1}^{n} e_i = 0 \\ < \mathbf{e}, \mathbf{x} > &= \sum_{i=1}^{n} X_i e_i = 0 \\ < \mathbf{e}, \widehat{\mathbf{Y}} > &= \sum_{i=1}^{n} \hat{Y}_i e_i = 0 \end{aligned} \right.$$

Geometric Interpretation of L-S regression

Figure: Orthogonal projection of response vector **Y** onto the linear subspace of $\mathbb{R}^n$ generated by the columns of the design matrix **X**



$X\hat{\beta} = \beta_0 \cdot \vec{1}_n + \beta_1 \cdot \vec{x} \in col(x)$

Sums of Squares: Matrix Form

Error Sum of Squares

$$SSE = \sum_{i=1}^{n} e_i^2$$

can be expressed in matrix form:

$= \langle e, e \rangle$ inner product of itself

$$SSE = \mathbf{e}'\mathbf{e} = \mathbf{Y}'\underbrace{(\mathbf{I}_n - \mathbf{H})'(\mathbf{I}_n - \mathbf{H})}\mathbf{Y} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}.$$

(quadratic form)

- $\mathbf{I}_n - \mathbf{H}$ is a projection matrix. $\langle X \rangle^{\perp}$ linear subspace $\perp$ to the col space
- $df(SSE) = rank(\mathbf{I}_n - \mathbf{H}) = n - 2.$ $(I_n - H)w = w - Hw \perp col(X)$

Total Sum of Squares

$\frac{1}{n} \cdot Y^T \cdot J_n Y$

$//$

$$SSTO = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n(\overline{Y})^2$$

can be expressed in matrix form:

$$SSTO = \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}_n\mathbf{Y} = \mathbf{Y}'\left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right)\mathbf{Y}.$$

(HW3)

- $\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$ is a projection matrix.

rank : max linearly independent col

$$\underset{\substack{n\times n\\ \text{of } 1}}{\mathbf{J}_n} = \mathbf{1}_n\mathbf{1}_n' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

$J_n (1_n, 1_n \cdots 1_n)$

$\hookrightarrow$ rank $(J_n) = 1$

- $df(SSTO) = rank(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n) = n - 1.$

SSTO = SSE + SSR

$\downarrow$ $\quad$ $\downarrow$

$y^T(I_n - \frac{1}{n}J_n)y$ $\quad$ $y^T(I_n - H)y$ $\quad \Rightarrow \quad y^T \cdot (H - \frac{1}{n}J_n)y$

Regression Sum of Squares

$$SSR = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2$$

can be expressed in matrix form:

$$\begin{aligned} SSR &= \left(\widehat{\mathbf{Y}} - \overline{\mathbf{Y}}\right)'\left(\widehat{\mathbf{Y}} - \overline{\mathbf{Y}}\right), \\ &= \mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}_n\right)'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}_n\right)\mathbf{Y} \\ &= \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)\mathbf{Y} \end{aligned}$$

$$\overline{\mathbf{Y}} = \frac{1}{n}\mathbf{J}_n\mathbf{Y} = \underbrace{\begin{pmatrix} \bar{y} \\ \bar{y} \\ \bar{y} \\ \vdots \end{pmatrix}}_{n\times 1}$$

- $\mathbf{H} - \frac{1}{n}\mathbf{J}_n$ is a projection matrix. HW3
- $df(SSR) = rank(\mathbf{H} - \frac{1}{n}\mathbf{J}_n) = 1.$ for simple regression

Expectation of SSE

$$
\begin{aligned}
E(SSE) &= E\left(\mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}\right) = E\left(Tr((\mathbf{I}_n - \mathbf{H})\mathbf{Y}\mathbf{Y}')\right) \\
&= Tr\left((\mathbf{I}_n - \mathbf{H})E(\mathbf{Y}\mathbf{Y}')\right) \\
&= Tr\left((\mathbf{I}_n - \mathbf{H})(\sigma^2\mathbf{I}_n + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')\right) \\
&= \sigma^2 Tr\left(\mathbf{I}_n - \mathbf{H}\right) + Tr\left((\mathbf{I}_n - \mathbf{H})\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'\right) \\
&= \underbrace{\overset{\text{df. of SSE}}{(n-2)}\sigma^2}
\end{aligned}
$$

$\Rightarrow MSE = \frac{SSE}{n-2}$ (an unbiased estimator of σ^2)

The last equality is because $Tr\left(\mathbf{I}_n - \mathbf{H}\right) = n - 2$ and $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0}$.

$$SSE = y^T (I_n - H) y$$

$$E(SSE) = E(Tr(SSE))$$

$$\overset{\text{def of SSE}}{=} E(Tr(y^T (I_n - H) y))$$

$$\overset{②}{=} E(Tr(\underbrace{(I_n - H)}_{\text{non-random}} y \cdot y^T))$$

$$\overset{\text{linearity}}{=} Tr[E((I_n - H) y \cdot y^T)]$$

$$\overset{①}{=} Tr[(I_n - H) \cdot (\sigma^2 I_n + x\beta \cdot \beta^T \cdot x^T)]$$

$$= Tr(\sigma^2 (I_n - H)) + Tr(\underbrace{(I_n - H) x}_{=0} \cdot \beta \cdot \beta^T \cdot x^T)$$

$$= \sigma^2 \cdot Tr(I_n - H)$$

$$= \sigma^2 \cdot [Tr(I_n) - Tr(H)]$$

$$= \sigma^2 \cdot (n - Tr(H))$$

$$= \sigma^2 \cdot (n - 2)$$

① $y = x\beta + \varepsilon$

$Ey = x\beta$, $\sigma^2\{y\} = \sigma^2 \cdot I_n$ } simple regression model

$$\underbrace{E(yy^T)}_{n\times n} = \sigma^2 \{y\} + (Ey) \cdot (Ey)^T$$

$$= \sigma^2 \times I_n + (x\beta) \times (x\beta)^T$$

$$= \sigma^2 \times I_n + x\beta \cdot \beta^T \cdot x^T$$

② property of Trace operator:

$tr(AB) = tr(BA)$ [sum of diag]

③ If $a \in \mathbb{R}$ is a real #, then $tr(a) = a$

$$Tr(H) = Tr(x(x^Tx)^{-1}x^T)$$

$$= Tr((x^Tx)^{-1} \cdot x^T x)$$

$$= Tr(I_2)$$

$$= 2$$

If $z \in \mathbb{R}^n$ is r.v., then

$$\underbrace{E(zz^T)}_{n\times n} = \underbrace{\sigma^2\{z\}}_{\text{var.cov matrix of } z} + (Ez) \cdot (Ez)^T$$

$\overset{\text{generalize}}{\Longleftarrow}$

z is a r.v., then

$$E(z^2) = var(z) + (E(z))^2$$

by def. of var.

optional

Confidence Interval for σ^2

Under the Normal error model:

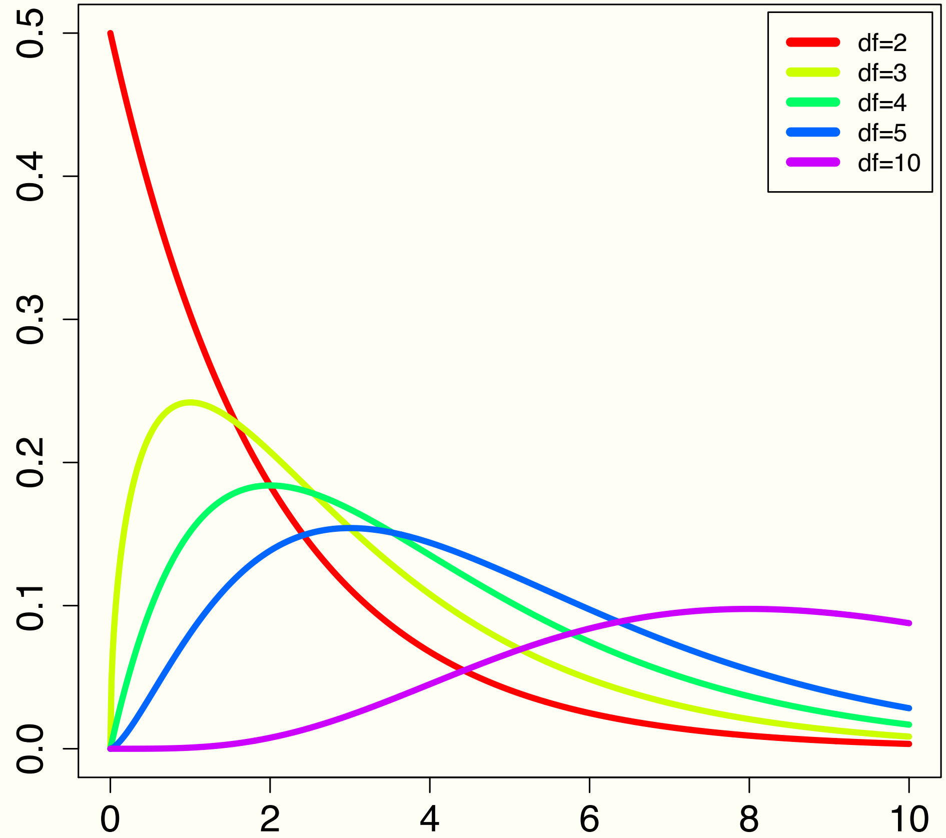
- $SSE \sim \sigma^2 \chi^2_{(n-2)}$
- Pivotal quantity:

$$\frac{SSE}{\sigma^2} \sim \chi^2_{(n-2)}$$

- $(1-\alpha)100\%$-confidence interval for σ^2:

$$\left[\frac{SSE}{\chi^2(1-\alpha/2; n-2)}, \frac{SSE}{\chi^2(\alpha/2; n-2)}\right]$$

Probability Density Curves of χ^2 Distributions



Properties of Projection Matrices

For projection matrix :

rank = trace

Optional material.

- Eigen-decomposition: $Q\Lambda Q^T$, where Q is an orthogonal matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues.

Idempotent

- Eigenvalues are either 1 or 0.

= rank

- The number of nonzero eigenvalues equals trace of the matrix equals the rank.
- For example. In simple linear regression:

$$\mathit{rank}(\mathbf{H}) \underset{=}{=tr(H)} 2, \quad \mathit{rank}(\mathbf{I}_n - \mathbf{H}) \underset{=}{=tr(I_n - H)} n - 2$$

Sampling Distribution of *SSE* under Normal Error Model

Optional material (Cont'd).

- $\mathbf{I}_n - \mathbf{H}$ is a projection matrix with rank $n-2 \Longrightarrow$

$$\mathbf{I}_n - \mathbf{H} = \mathbf{Q}^T \Lambda \mathbf{Q},$$

  where $\Lambda = \mathrm{diag}\{1, \cdots, 1, 0, 0\}$ and $\mathbf{Q}$ is an orthogonal matrix.

- $(\mathbf{I}_n - \mathbf{H})\mathbf{X} = \mathbf{0} \Longrightarrow$

$$\mathbf{e} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y} = (\mathbf{I}_n - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = (\mathbf{I}_n - \mathbf{H})\boldsymbol{\epsilon}$$

Optional material (Cont'd).

- $SSE = \mathbf{e}^T\mathbf{e} = \boldsymbol{\epsilon}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\epsilon} = (\mathbf{Q}\boldsymbol{\epsilon})^T\Lambda(\mathbf{Q}\boldsymbol{\epsilon})$
- Let $\mathbf{z} = \mathbf{Q}\boldsymbol{\epsilon}$, then

$$SSE = \sum_{i=1}^{n-2} z_i^2$$

- Moreover

$$\mathbf{E}(\mathbf{z}) = \mathbf{Q}\mathbf{E}\{\boldsymbol{\epsilon}\} = \mathbf{0}, \quad \boldsymbol{\sigma}^2\{\mathbf{z}\} = \mathbf{Q}\boldsymbol{\sigma}^2\{\boldsymbol{\epsilon}\}\mathbf{Q}^T = \sigma^2\mathbf{Q}\mathbf{Q}^T = \sigma^2\mathbf{I}_n$$

  So under Normal error model, z_is are i.i.d. $N(0, \sigma^2)$.

- Thus $SSE \sim \sigma^2\chi^2_{(n-2)}$.