Multiple Linear Regression

Outline:

- Model setup & LS estimates of matrix multiplication
- Gauss-Markov for MLR (skipped)
- Fitted values & residuals / Estimate σ^2

1 Model set up

Start w/ simple linear reg.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{for } i = 1, \dots, n$$

$\varepsilon_i \sim N(0, \sigma^2)$

In other words:

$$y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1$$
$$y_2 = \beta_0 + \beta_1 x_2 + \varepsilon_2$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 x_n + \varepsilon_n$$

Let $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$,

$$Y = \underset{n\times 2}{X}\,\underset{2\times 1}{\beta} + \varepsilon$$

$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$,

$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$, &

$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$.

Assumptions:

① $E(\varepsilon) = 0$

$Var(x) = E[(X-\mu)^2]$

② $Var(\varepsilon) = \sigma^2 I_n$

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix}$$

$$\underbrace{Var(X)}_{n\times n} = E\Big[\underbrace{(X-\mu)}_{n\times 1}\underbrace{(X-\mu)^T}_{1\times n}\Big]$$

$$= \begin{bmatrix} Var(X_1) & Cor(X_1, X_2) \\ Cor(X_1, X_2) & Var(X_2) \end{bmatrix}$$

"variance-covariance matrix"

If $\varepsilon_i \perp\!\!\!\perp \varepsilon_j \ \forall\, i \neq j$, then

$$Var(\varepsilon) = \begin{bmatrix} Var(\varepsilon_1) & 0 & 0 & \cdots & 0 \\ 0 & Var(\varepsilon_2) & & & \vdots \\ \vdots & & \ddots & & \\ & & & \ddots & 0 \\ 0 & \text{—} & \cdots & 0 & Var(\varepsilon_n) \end{bmatrix}$$

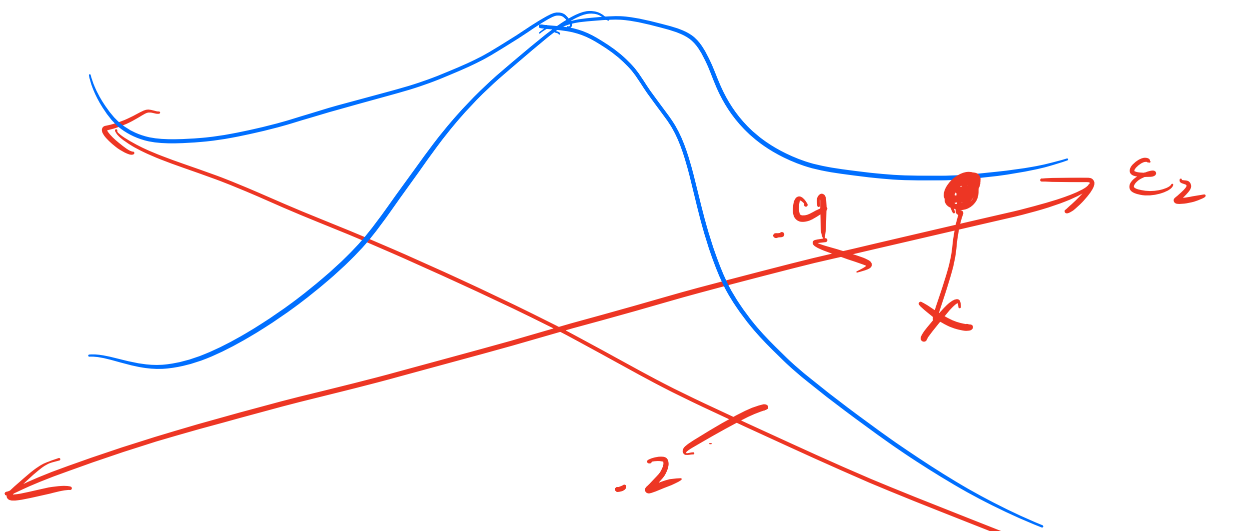
$Var(\varepsilon_i) = \sigma^2 \ \forall\, i = 1, \ldots, n$

$$\mathrm{Var}(\varepsilon) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2 \end{bmatrix} = \underline{\underline{\sigma^2 I_n}}$$

---

(3) $\varepsilon \sim \underline{\underline{N}}(0, \sigma^2 I_n)$

"multivariate normal"



pdf of multivariate Normal:
w/ mean 0 & covariance $\sigma^2 I_n$

$$f_\varepsilon(\varepsilon_1, \ldots, \varepsilon_n) = \prod_{i=1}^{n} f_{\varepsilon_i}(\varepsilon_i)$$

.2 .6

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\varepsilon_i - \mu)^2 / 2\sigma^2}$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\sum_{i=1}^{n} (\varepsilon_i - \mu)^2 / 2\sigma^2}$$

$\boxed{2}$ Parameter Estimation under Matrix Setup

Idea: minimize least squares criterion

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Equivalently:

$$Q(\beta) = \|Y - X\beta\|^2$$

$$= \underset{1 \times n}{(Y - X\beta)^T} \underset{n \times 1}{(Y - X\beta)}$$

We want to minimize this w.r.t. β.

Refresher: Matrix Differentiation

① $\dfrac{\partial(\beta^T A)}{\partial \beta} = A$ ② $\dfrac{\partial(\beta^T A \beta)}{\partial \beta} = 2A\beta$

⊛ Think of univariate versions & then make dims work

Minimize

$$Q(\beta) = (Y - X\beta)^T (Y - X\beta)$$

$Y^T(X\beta) = (Y^T(X\beta))^T = (X\beta)^T Y$
$= \beta^T X^T Y$

$$= Y^T Y - Y^T X\beta - (X\beta)^T Y + (X\beta)^T (X\beta)$$

$$= Y^T Y - 2\underbrace{\beta^T X^T Y} + \beta^T X^T X \beta$$

$$\frac{\partial Q}{\partial \beta} = 0 - 2X^T Y + 2(X^T X)\beta \overset{!}{=} 0$$

$$\Rightarrow \cancel{2}(X^T X)\beta = \cancel{2}X^T Y$$

$$\Rightarrow \boxed{\hat{\beta} = (X^T X)^{-1} X^T Y}$$

OLS

"Normal Eq."

For SLR the normal eq. gives:

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \& \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \text{So}$$

$$X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_i x_i^2 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{n\sum_i x_i^2 - (n\bar{x})^2} \begin{bmatrix} \sum_i x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

$$(X^T X)(X^T X)^{-1} = \frac{1}{n\sum_i x_i^2 - (n\bar{x})^2} \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} \sum_i x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$X^T Y = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ \sum_i x_i y_i \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{n\sum_i x_i^2 - (n\bar{x})^2} \begin{bmatrix} \sum_i x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix} \begin{bmatrix} n\bar{y} \\ \sum_i x_i y_i \end{bmatrix}$$

$$= \frac{1}{nSSX} \begin{bmatrix} n\bar{y}\sum_i x_i^2 - n\bar{x}\sum_i x_i y_i \\ -n^2\bar{x}\bar{y} + n\sum_i x_i y_i \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\bar{y}\sum_i x_i^2 - \bar{x}\sum_i x_i y_i}{SSX} \\ \\ \dfrac{\sum_i x_i y_i - n\bar{x}\bar{y}}{SSX} \end{bmatrix} = \begin{bmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \\ \dfrac{SSXY}{SSX} \end{bmatrix}$$

$$\frac{SSXY}{SSX} = \hat{\beta}_1$$

$$\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

Num:

$$\frac{\bar{y}\sum_i x_i^2}{SSX} - \frac{\bar{x}\sum_i x_i y_i}{SSX} - \frac{n\bar{x}^2\bar{y}}{SSX} + \frac{n\bar{x}^2\bar{y}}{SSX}$$

$$\bar{y}\cancel{\frac{(\sum_i x_i^2 - n\bar{x}^2)}{SSX}} - \frac{\bar{x}(\sum_i x_i y_i - n\bar{x}\bar{y})}{SSX}$$

$$= \bar{y} - \bar{x}\hat{\beta}_1$$

Concavity Check:

$$\frac{\partial^2 Q}{\partial \beta^2} = \underline{\underline{2X^TX}}$$

A matrix $\overset{(p\times p)}{M}$ is "positive definite" iff

$$\forall x \in \mathbb{R}^p, \quad x^T M x > 0.$$

& "positive semidefinite" iff

$$\forall x \in \mathbb{R}^p, \quad x^T M x \geq 0.$$

Is $\frac{\partial^2 Q}{\partial \beta^2}$ positive semidefinite?

For any other vector $v \in \mathbb{R}^n$

$$\begin{aligned} v^T(2X^TX)v &= 2 v^T X^T X v \\ &= 2 (Xv)^T (Xv) \\ &= 2 \|Xv\|^2 \geq 0 \end{aligned}$$

$\therefore$ $\partial^2 Q / \partial \beta^2$ is psd. $\Rightarrow \hat{\beta} = (X^TX)^{-1}X^Ty$ is a minimizer.

The matrix setup allows us to consider more than one predictor!

Ex:

Multiple Lin. Reg Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i$$

has the matrix form

$Y = X\beta + \varepsilon$ where

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{(p-1)1} \\ 1 & x_{12} & \cdots & & \vdots \\ \vdots & \vdots & & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{(p-1)n} \end{bmatrix} \quad \& \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

$\therefore$ Solution for MLR is just:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

What about a matrix expression for fitted values?

$$\hat{Y} = X\hat{\beta}$$

$$= X(X^TX)^{-1}X^TY$$

$$= \underset{n\times n}{H}\,\underset{n\times 1}{Y}$$

Hat matrix:

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

Claim:

H is idempotent $\iff H^2 = H$.

Pf:

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

$$= X\,I\,(X^TX)^{-1}X^T$$

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

Residuals:

$$e_i = y_i - \hat{y}_i$$

$$e = Y - \hat{Y} = Y - HY = (I-H)Y$$

Claim: $I-H$ is idempotent.

Pf: $(I-H)^2 = (I-H)(I-H)$

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

$$= I - H - \cancel{H + H} = I - H$$

How to estimate σ^2?

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

$$= ((I-H)Y)^T [(I-H)Y]$$

$$= Y^T (I-H)^T (I-H) Y$$

$$= Y^T (I-H)(I-H) Y$$

$$= Y^T (I-H) Y$$

$$y \sim N(X\beta, \sigma^2 I_n)$$

$$\Rightarrow (I-H) \quad Y \sim N((I-H)X\beta, \sigma^2 (I-H))$$

$$\Rightarrow \frac{(I-H)y}{\sigma} \sim N(0, (I-H))$$

Idea:

$$\underbrace{SSE}_{\sigma^2} = \left\| \frac{(I-H)Y}{\sigma} \right\|^2 \sim \chi^2_{df}$$

$$df = \text{rank}(I-H) = n-p$$

Then:

$$E\left(\frac{SSE}{\sigma^2}\right) = n-p$$

$$\Rightarrow E\left(\underbrace{\frac{SSE}{n-p}}_{\hat{\sigma}^2}\right) = \sigma^2$$

$\hat{\sigma}^2 = \frac{SSE}{n-p} = \frac{e^Te}{n-p}$ is unbiased for σ^2.

Symmetry of H & $(I-H)$:

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

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

$$= X\ [(X^TX)^T]^{-1} X^T$$

$$= X\ [X^TX]^{-1} X^T = H \Rightarrow H \text{ symmetric.}$$

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