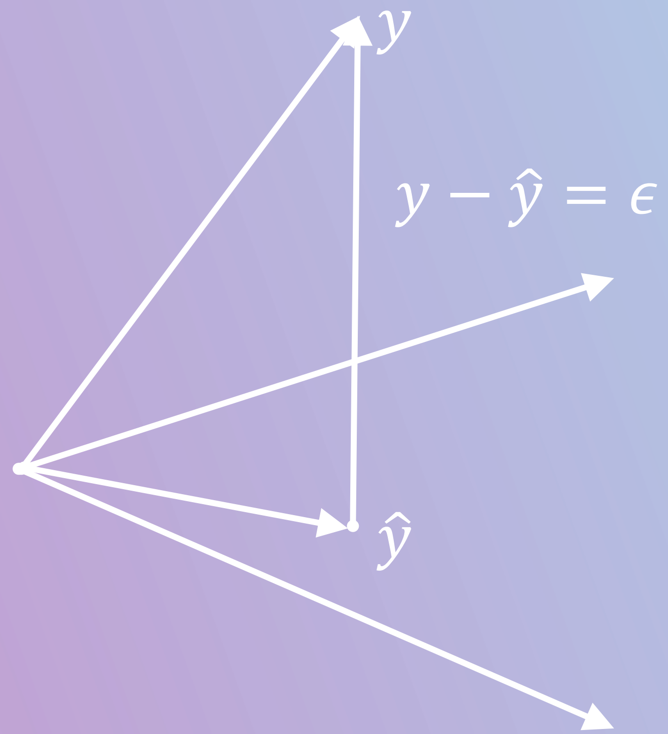
YONSEI ESC

HD, Statduck

Week2 Regression



오늘의 이야기

< 1부 >

1. 서론: 통계적 관점과 기하적 관점
   - 상관계수와 정사영 예시
   - 그램슈미트 과정과 QR 분해

2. 선형회귀 - 통계

3. 선형회귀 - 기하

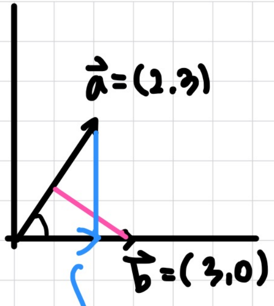
< 2부 >

1. 베타
   - 식 리뷰
   - Gauss Markov
   - Always good!

2. 연속적인 직교

1부 — 서론: 통계적 관점과 기하학적 관점

상관계수와 정사영 예시



$proj_b a = \frac{\vec{a}\cdot\vec{b}}{|\vec{b}|}$

$\Rightarrow \vec{a}\cdot\vec{b} = |\vec{a}||\vec{b}|\cos\theta \longrightarrow \cos\theta = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}$

$= |\vec{a}|\cos\theta \times |\vec{b}|$

: 벡터 a의 변화를 벡터 b가 얼만큼 설명해줄 수 있는가?

물론, $proj_a b = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}|}$, : 벡터 b의 변화를 벡터 a가 얼만큼 설명?

$\Rightarrow \frac{1}{|\vec{a}|}$: a가 b를 설명하는 정도 $\frac{1}{|\vec{b}|}$: b가 a를 설명하는 정도

$\vec{a}\cdot\vec{b}$: $\vec{a}$와 $\vec{b}$의 관계

$\cos\theta = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\cdot|\vec{b}|}$: $\vec{a}$와 $\vec{b}$가 서로를 설명하는 양

1부 — 서론:통계적관점과 기하적 관점

상관계수와 정사영 예시

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \overline{X}}{s_{\overline{X}}}\right)\left(\frac{y_i - \overline{Y}}{s_{\overline{Y}}}\right)$$

$$= \frac{1}{n-1}\frac{\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})}{\left\{\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(x_i - \overline{X})^2\right\}^{\frac{1}{2}}\left\{\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(y_i - \overline{Y})^2\right\}^{\frac{1}{2}}}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})}{\left\{\sum_{i=1}^{n}(x_i - \overline{X})^2\right\}^{\frac{1}{2}}\left\{\sum_{i=1}^{n}(y_i - \overline{Y})^2\right\}^{\frac{1}{2}}}$$

: 피어슨 상관계수

$$\begin{cases} a_i = x_i - \overline{X} \\ b_i = y_i - \overline{Y} \end{cases}$$

$$= \frac{\sum_{i=1}^{n} a_i b_i}{\left\{\left(\sum_{i=1}^{n} a_i a_i\right)\left(\sum_{i=1}^{n} b_i b_i\right)\right\}^{\frac{1}{2}}} = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}$$

: 위의 $\cos\theta$와 표현하는 것이 같아짐

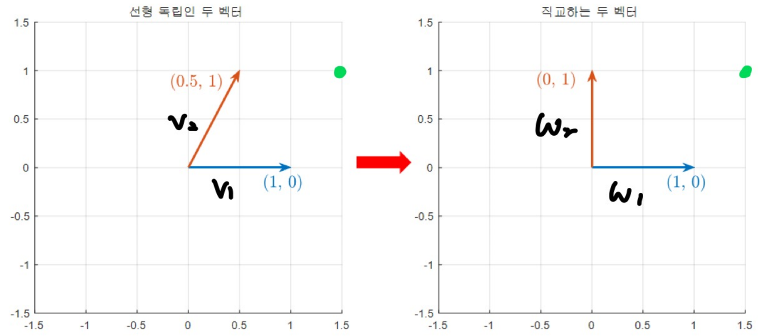
$= $ 상관계수 r

: $x_i - \overline{X}$ 와 $y_i - \overline{Y}$ 가 서로를 설명하는 양

$$\begin{bmatrix} -1 \le \cos\theta \le 1 \\ -1 \le r \le 1 \end{bmatrix}$$

상관계수 r 과 $\cos\theta$는 본질적으로 비슷한 의미를 공유하고 있다.

그람슈미트 과정과 QR 분해



$v = v_1 + v_2$
$= (1.5, 1)$

$v = \{(1.5,1)\cdot(1,0)\}\,w_1$
$+ \{(1.5,1)\cdot(0,1)\}\,w_2$
$= 1.5\,w_1 + w_2$

선형 독립인 두 벡터 v_1, v_2

→ $v \in \mathbb{R}^2$ 에 대해

$v = x_1 v_1 + x_2 v_2$, $x_1, x_2 \in \mathbb{R}$ 로 표현가능

: v는 v_1과 v_2의 선형 결합

만약, 직교하는 두 벡터 w_1, w_2 라면?

$v = (v\cdot w_1)w_1 + (v\cdot w_2)w_2$ 로 표현가능,

즉 계수를 직교기저와의
내적을 통해 쉽게 찾을 수 있다.

그램슈미트 과정과 QR 분해

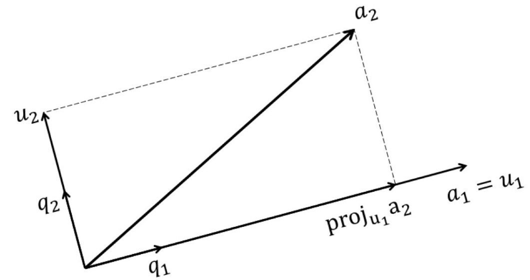
Gram-Schmit 과정

벡터공간 V, 일차 독립인 벡터 $\{a_1, \cdots, a_k\}$

$$\rightsquigarrow \begin{aligned} u_1 &= a_1 \\ u_2 &= a_2 - proj_{u_1}(a_2) \\ u_3 &= a_3 - proj_{u_1}(a_3) - proj_{u_2}(a_3) \\ &\vdots \\ u_k &= a_k - \sum_{j=1}^{k-1} proj_{u_j}(a_k) \end{aligned}$$

$\Rightarrow \{u_1, \cdots, u_k\}$: 직교적

$q_i = \dfrac{u_i}{\|u_i\|} \Rightarrow$ 벡터공간 V의 정규직교기저 $\{q_1, \cdots, q_k\}$



QR 분해

: 정규직교기저 벡터 이용해 행렬 분해

Q : 정규직교기저 $\{q_1, \cdots, q_n\}$ 모은 것

$\Rightarrow A = QR$

$$\begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix} \begin{bmatrix} a_1 q_1 & \cdots & a_n q_1 \\ \vdots & \ddots & \\ a_1 q_n & & a_n q_n \end{bmatrix}$$

: 계산해보면 성립함 (인정)

$*\ a_i \cdot q_j = 0$, if $i < j$

: q_j에는 $1, \cdots, j-1$ 성분에 대한 정보가 제거됨.

따라서, 위의 식은..

$$\begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix} \begin{bmatrix} a_1 q_1 & \cdots & a_n q_1 \\ & \ddots & \vdots \\ 0 & & a_n q_n \end{bmatrix}$$

정규직교기저행렬과 상삼각행렬로 분해 가능

f가 선형함수이면
선형 회귀분석

$$Y = f\left(X_1, \ldots, X_p\right) + \varepsilon$$

오차항(함수적 관계와는 관련이 없음)

회귀분석: 관심대상 Y와 X를 설명할 만한 정보들 $X_1, \ldots, X_p$ 의 함수적 관계인 $f(\cdot)$를 찾는 것

$$\mathrm{E}(Y|X)$$

선형회귀: 관심대상 Y를 설명할만한 정보들이 주어졌을 때, 관심대상에 대한 정보들의 영향을 추정(Beta)하여 관심대상 Y의 평균을 구하는 것, 즉 조건부 기댓값!

관심 대상: Y
주어진 정보: $X_1, \ldots, X_p$
영향력: $\beta_1, \ldots \beta_p$
특정조건 하에서 관심 대상의 평균: $E(Y|X_1 = x_1, \ldots, X_p = x_p, \beta_1, \ldots, \beta_p)$

$$\text{주택가격} = 2 \cdot \text{방의수} - 0.7 \cdot \text{지하철역과의 거리} + 10{,}000{,}000$$

위 모형이 맞다면 방의 수가 2개고 지하철역과의 거리가 1.2km이면 주택가격은 정확히 f(2,1.2)일까?

아니다.

그렇다면 주택가격을 나타내는 모든 점들은 직선 상에 모조리 위치해야한다.
이러면 모형이 우리가 탐구하는 생성 매커니즘을 다 설명하고 있는 것? 근데 그런 일은 없다.
오차의 존재 이유를 살펴봐야 한다.

주택가격은 두 예측변수 이외의 변수로부터 영향을 받는다.(시공년도)
예측변수 이외의 변수들의 영향이 없다 하더라도 판매량의 값에는
운 혹은 우연과 같은 어느 정도의 Randomness가 존재한다.

오차항: 예측변수 외 변수의 영향과 Randomness 반영해줌.

1부 — 선형회귀: 통계

$$\varepsilon \sim iid\ N(0, \sigma^2)$$

정규성

잔차가 정규분포가 되도록 회귀모형을 만들어야 모형이 데이터에 잘 적합한 것이다.

독립성

신뢰할 수 있는 조건부 평균은 데이터들이 특정 조건에 편향되지 않고 독립적으로 공평하게 수집되어야 한다. 즉 오차들이 서로 영향을 주면 안된다.

등분산성

조건별로 취합된 데이터들의 분산이 비슷해야 한다.

오차항의 **평균은 0**이어야 하나?

오차항들은 **독립**이어야 하나?

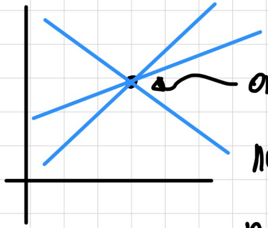
오차항의 **정규성 및 등분산성 가정**은 왜 하는건가?

예측변수의 개수는 $p-1$개, 회귀계수의 개수는 p개, 자료의 개수는 n개

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{1,1} + \cdots + \beta_{p-1} x_{1,p-1} + \varepsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_{2,1} + \cdots + \beta_{p-1} x_{2,p-1} + \varepsilon_2 \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_{n,1} + \cdots + \beta_{p-1} x_{n,p-1} + \varepsilon_n
\end{aligned}
$$

$$
\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p-1} \\ 1 & x_{2,1} & \cdots & x_{2,p-1} \\ 1 & x_{3,1} & \cdots & x_{3,p-1} \\ & & \ddots & \\ 1 & x_{n,1} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{pmatrix}
$$

극단적인 경우로 $n=1$ 이면 관측치가 하나다. 이때 $y = \beta_0 + \beta_1 X + \varepsilon$ 추정하면
점을 지나는 직선은 셀 수 없이 많고,
모든 line의 SSE를 0으로 만듦. 3차원 공간에서 생각해도 셀수없음



only one obs

no unique solution when p(unknown) $> n$(obs)

$n \gg p$: Machine learning model assumes

차원의 저주

데이터 수집⇑ as $p\uparrow$

or

So 차원축소 필요

p >> n 일 때의 주된 문제점은 우리가 사용할 머신러닝 기법의 모델들이 train set에 대해 overfitting 하게 된다는 것.
데이터가 적기 때문에 일반화하기 어렵고 그로 인해 통계적인 오차들도 train set의 정보라고 판단하기 쉽다.
그래서 이제 대표적인 해결 방법으로 회귀나 분류 문제에서는 변수 선택, pca, regularization 등의 방법을 사용한다.

Vector Derivatives Rule 3

Let $\alpha = \mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{x}$ where $\mathbf{y} \in \mathbb{R}^{m\times 1}$, $\mathbf{x} \in \mathbb{R}^{n\times 1}$, and $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then $\dfrac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}^{\mathbf{T}}\mathbf{A}$ and $\dfrac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}$

($\mathbf{A}$ is constant matrix.)

Vector Derivatives Rule 4 (Quadratic formula)

Let $\alpha = \mathbf{x}^{\mathbf{T}}\mathbf{A}\mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^{n\times 1}$, and $\mathbf{A} \in \mathbb{R}^{n\times n}$. Then $\dfrac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^{\mathbf{T}}(\mathbf{A} + \mathbf{A}^{\mathbf{T}})$. ($\mathbf{A}$ is constant matrix.)

Example: Least Sqaures Estimators

- For a target vector $\mathbf{t}$ and a design matrix $\mathbf{\Phi}$, LSE $\mathbf{w}$ can be obatained by;

$$\mathbf{w} = \arg\min_{\mathbf{w}} \|\mathbf{t} - \mathbf{\Phi}\mathbf{w}\|^2 = (\mathbf{t} - \mathbf{\Phi}\mathbf{w})^T(\mathbf{t} - \mathbf{\Phi}\mathbf{w})$$

$$\frac{\partial}{\partial \mathbf{w}}(\mathbf{t}^{\mathbf{T}}\mathbf{t} - 2\mathbf{t}^{\mathbf{T}}\mathbf{\Phi}\mathbf{w} + \mathbf{w}^{\mathbf{T}}\mathbf{\Phi}^{\mathbf{T}}\mathbf{\Phi}\mathbf{w}) =^{set} 0$$

$$\therefore \mathbf{w}_{OLS} = (\mathbf{\Phi}^{\mathbf{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathbf{T}}\mathbf{t}$$

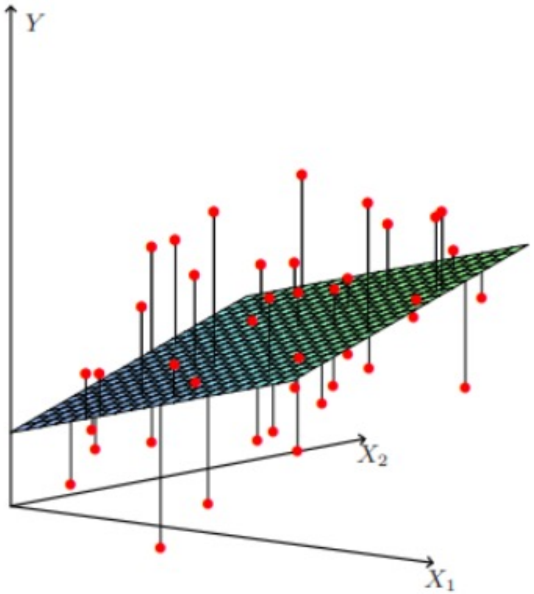


Figure: RSS when $X \in \mathbb{R}^2$

$$f(X) = \beta_0 + X_1\beta_1 + X_2\beta_2$$
$$RSS(\beta) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$
$$\frac{\partial RSS}{\partial \beta} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta)$$
$$\frac{\partial^2 RSS}{\partial \beta \partial \beta^T} = -2\mathbf{X}^T\mathbf{X}$$

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) = 0$$
$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$
$$\hat{y} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{H}\mathbf{y}$$

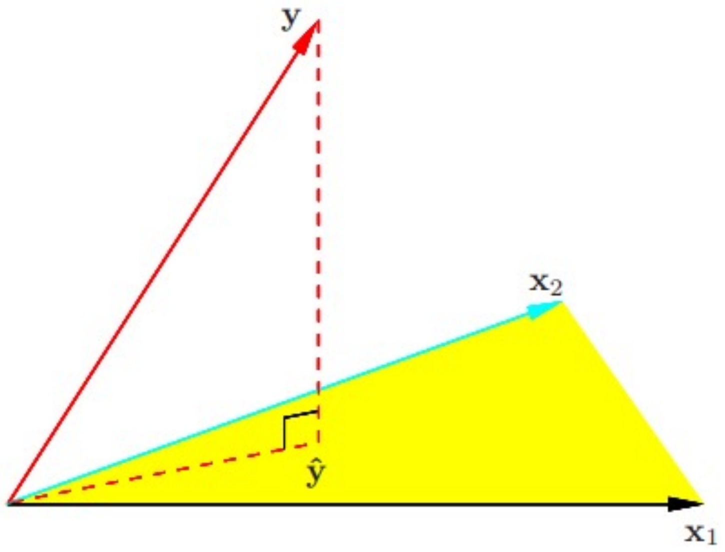


Figure: Linear regression in Geometry

$$\begin{aligned}\hat{\mathbf{Y}} &= \mathbf{Xb} = \underbrace{\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T}_{\mathbf{H}}\mathbf{Y} = \mathbf{HY}\\ \mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}}\\ &= (\mathbf{I}-\mathbf{H})\mathbf{Y}\\ \mathbf{H} &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \quad \text{is called Hat Matrix}\end{aligned}$$

$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is a *Projection* matrix

$$\begin{aligned}\mathbf{H}^T &= \mathbf{H}: && \text{Symmetric}\\ \mathbf{HH} &= \mathbf{H}: && \text{Idempotent}\\ \mathbf{HX} &= \mathbf{X}\\ \mathbf{X}^T\mathbf{H} &= \mathbf{X}^T\\ (\mathbf{I}-\mathbf{H})^2 &= \mathbf{I}-\mathbf{H}\\ (\mathbf{I}-\mathbf{H})\mathbf{H} &= \mathbf{H}(\mathbf{I}-\mathbf{H}) = \mathbf{0}\end{aligned}$$

- $\sum_{i=1}^n e_i^2 =$ Minimum of $Q(\beta)$
- $\sum_{i=1}^n e_i = 0$
- $\sum_{i=1}^n X_{ij}e_i = 0$ for all $j = 1, \ldots, p-1$.
- $\sum_{i=1}^n Y_i = \sum_{i=1}^n \hat{Y}_i$
- $\sum_{i=1}^n \hat{Y}_i e_i = 0$

$$Y = f(X) + \varepsilon = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1} + \varepsilon$$

$$\varepsilon \sim iid\ N(0, \sigma^2)$$

우리는 반응변수(Y)가 어떤 매커니즘에 의해 생성되는지 알 수 없다.

우리가 가지고 있는건 n개의 관측 데이터 뿐이고, 이를 이용해 회귀계수들의 참값을 추정해보자!

~ 1부 마무리 ~

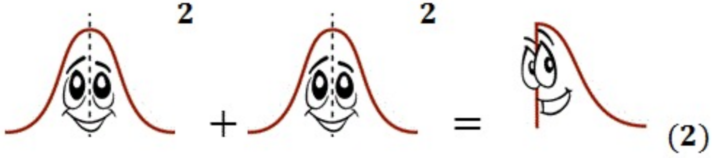
우리, 쉬어가요

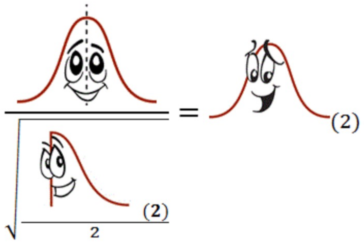
Beta Estimation

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\beta + \epsilon)$$
$$\hat{\beta} \sim N(\beta, (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2)$$

$$\hat{\sigma}^2 = \frac{1}{N-p-1}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$
$$(N-p-1)\hat{\sigma}^2 \sim \sigma^2\chi^2_{N-p-1}$$

$$z_j = \frac{\hat{\beta}_j}{\hat{sd}(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{v_j}} \sim t(df) \quad s.t.\ N-p-1 = df$$





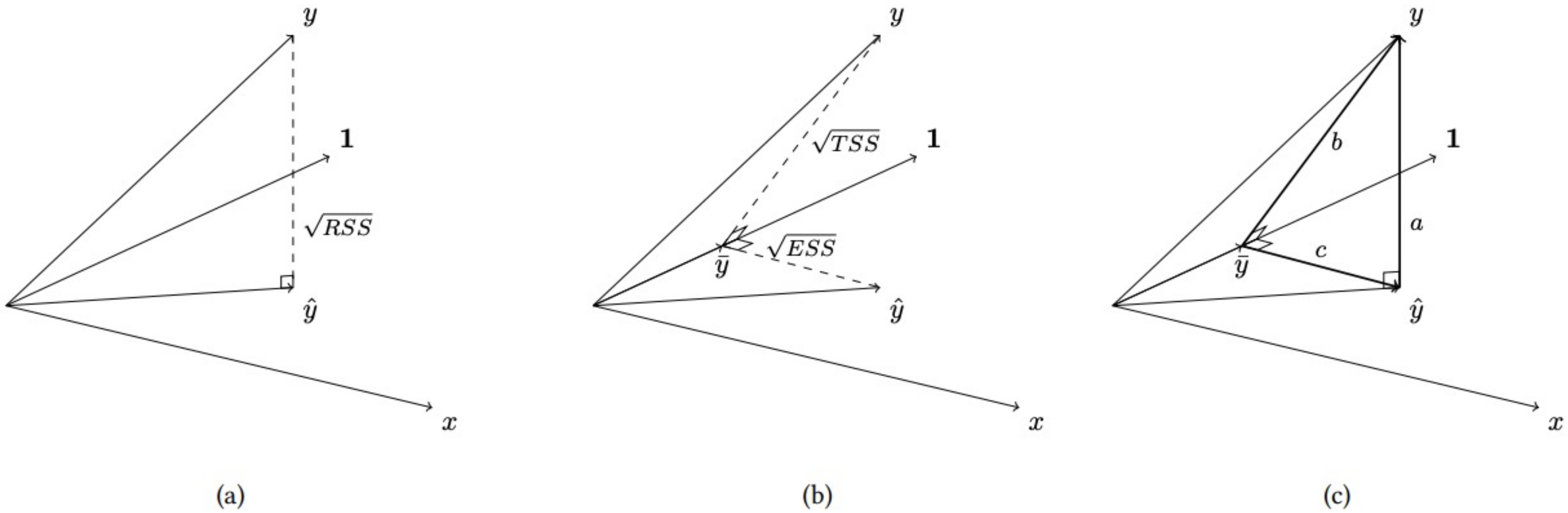


Figure: Linear regression in Geometry

$$F = \frac{among\ group\ var}{within\ group\ var}$$

$$\frac{MSR}{MSE} = \frac{(RSS_0 - RSS_1)/(p_1 - p_0)}{RSS_1/(N - p_1 - 1)}$$

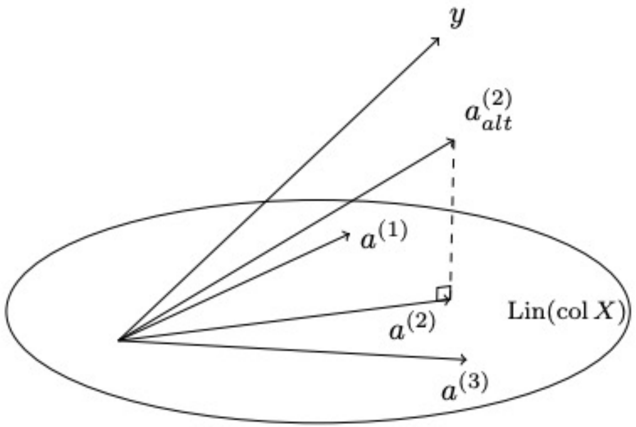


Figure: GM in Geometry

$$\tilde{\beta} = Cy,\ C = (X^T X)^{-1}X + D,\ D:\ K \times n\ matrix$$

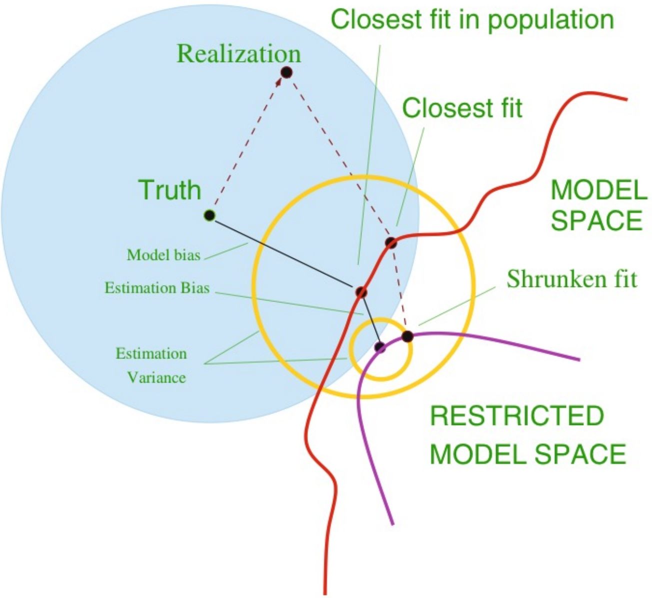
$$\begin{aligned}
\mathrm{E}\left[\tilde{\beta}\right] &= \mathrm{E}[Cy] \\
&= \mathrm{E}\left[\left((X'X)^{-1}X' + D\right)(X\beta + \varepsilon)\right] \\
&= \left((X'X)^{-1}X' + D\right)X\beta + \left((X'X)^{-1}X' + D\right)\mathrm{E}[\varepsilon] \\
&= \left((X'X)^{-1}X' + D\right)X\beta \qquad (\mathrm{E}[\varepsilon] = 0) \\
&= (X'X)^{-1}X'X\beta + DX\beta \\
&= (I_K + DX)\beta.
\end{aligned}$$

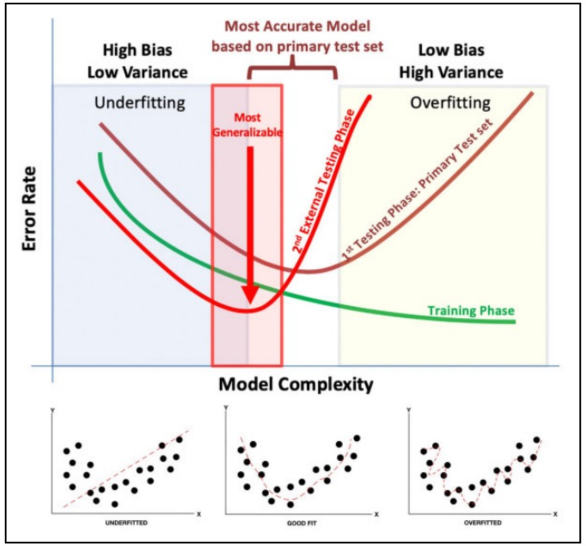
$$\begin{aligned}
\mathrm{Var}\left(\tilde{\beta}\right) &= \mathrm{Var}(Cy) \\
&= C\ \mathrm{Var}(y)C' \\
&= \sigma^2 CC' \\
&= \sigma^2\left((X'X)^{-1}X' + D\right)\left(X(X'X)^{-1} + D'\right) \\
&= \sigma^2\left((X'X)^{-1}X'X(X'X)^{-1} + (X'X)^{-1}X'D' + DX(X'X)^{-1} + DD'\right) \\
&= \sigma^2(X'X)^{-1} + \sigma^2(X'X)^{-1}(DX)' + \sigma^2 DX(X'X)^{-1} + \sigma^2 DD' \\
&= \sigma^2(X'X)^{-1} + \sigma^2 DD' \qquad (DX = 0) \\
&= \mathrm{Var}\left(\hat{\beta}\right) + \sigma^2 DD' \qquad \left(\sigma^2(X'X)^{-1} = \mathrm{Var}\left(\hat{\beta}\right)\right)
\end{aligned}$$

DD' is a positive semi-definite matrix.(∵ it is a symmetric matrix.) $\hat{\beta}^{LS}$ is MVUE(Minimum Variance Unbiased Estimator).

2부 — 베타: Always good?

No, least square method is not always the best.





$$
\begin{aligned}
Err(x_0) &= E[(Y - \hat{f}(x_0))^2 | X = x_0] \\
&= \sigma_\epsilon^2 + [E\hat{f}(x_0) - f(x_0)]^2 + E[\hat{f}(x_0) - E\hat{f}(x_0)]^2 \\
&= \sigma_\epsilon^2 + Bias^2(\hat{f}(x_0)) + Var(\hat{f}(x_0)) \\
&= Irreducible\ Error + Bias^2 + Variance
\end{aligned}
$$

Why use?

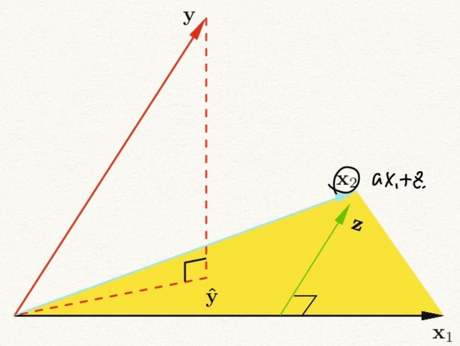


FIGURE 3.4. *Least squares regression by orthogonalization of the inputs. The vector $\mathbf{x}_2$ is regressed on the vector $\mathbf{x}_1$, leaving the residual vector $\mathbf{z}$. The regression of $\mathbf{y}$ on $\mathbf{z}$ gives the multiple regression coefficient of $\mathbf{x}_2$. Adding together the projections of $\mathbf{y}$ on each of $\mathbf{x}_1$ and $\mathbf{z}$ gives the least squares fit $\hat{\mathbf{y}}$.*

Algorithm 3.1 *Regression by Successive Orthogonalization.*

1. Initialize $\mathbf{z}_0 = \mathbf{x}_0 = \mathbf{1}$.
2. For $j = 1, 2, \ldots, p$

   Regress $\mathbf{x}_j$ on $\mathbf{z}_0, \mathbf{z}_1, \ldots, , \mathbf{z}_{j-1}$ to produce coefficients $\hat{\gamma}_{\ell j} = \langle \mathbf{z}_\ell, \mathbf{x}_j \rangle / \langle \mathbf{z}_\ell, \mathbf{z}_\ell \rangle$, $\ell = 0, \ldots, j-1$ and residual vector $\mathbf{z}_j = \mathbf{x}_j - \sum_{k=0}^{j-1} \hat{\gamma}_{kj} \mathbf{z}_k$.
3. Regress $\mathbf{y}$ on the residual $\mathbf{z}_p$ to give the estimate $\hat{\beta}_p$.

x_1 on z_0 $\rightarrow$ $x_1 = \gamma_{01} + z_1$

x_2 on z_0, z_1

$$Y = X\beta + \varepsilon$$
$$\hat{\beta} = (X^T X)^{-1} X^T y$$
$$\hat{\beta}_j = \frac{\langle \mathbf{x}_j, \mathbf{y} \rangle}{\langle \mathbf{x}_j, \mathbf{x}_j \rangle}, \quad when\ X = Orthogonal\ matrix$$

Matrix form

$$
\begin{aligned}
\mathbf{X} &= \mathbf{Z\Gamma} \\
&= \begin{bmatrix} z_{11} & \cdots & z_{p1} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{np} \end{bmatrix}
\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1p} \\ 0 & \gamma_{22} & \cdots & \gamma_{2p} \\ 0 & 0 & \cdots & \gamma_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_{pp} \end{bmatrix} \\
&= \mathbf{ZD}^{-1}\mathbf{D\Gamma} \\
&= \mathbf{QR} \\
&= \begin{bmatrix} q_1 & \cdots & q_p \end{bmatrix}
\begin{bmatrix} (x_1 \cdot q_1) & (x_2 \cdot q_1) & \cdots & (x_p \cdot q_1) \\ 0 & (x_2 \cdot q_2) & \cdots & (x_p \cdot q_2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (x_p \cdot q_p) \end{bmatrix}
\end{aligned}
$$

과제 설명

https://github.com/YonseiESC/ESC-21SUMMER/tree/main/week1/HW

[1번 문제] ISL 3.5

5. Consider the fitted values that result from performing linear regression without an intercept. In this setting, the ith fitted value takes the form

$$\hat{y}_i = x_i\hat{\beta},$$

where

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i y_i\right) / \left(\sum_{i'=1}^{n} x_{i'}^2\right). \tag{3.38}$$

Show that we can write

$$\hat{y}_i = \sum_{i'=1}^{n} a_{i'} y_{i'}.$$

What is $a_{i'}$?

[2번 문제] ESL 3.4

Ex. 3.4 Show how the vector of least squares coefficients can be obtained from a single pass of the Gram–Schmidt procedure (Algorithm 3.1). Represent your solution in terms of the QR decomposition of $\mathbf{X}$.

[3번 문제] 코딩

Numpy만 이용해서 선형회귀 코드짜기

깃헙에 주어진 데이터셋

회귀분석,

즐거운 이야기

마침.