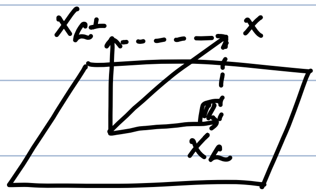
* Orthogonal projection

- $\mathcal{L} \in \mathbb{R}^n$을 따를 때, $\mathcal{L}^\perp = \{x \in \mathbb{R}^n : x^T y = 0 \text{ for all } y \in \mathcal{L}\}$
- $\mathcal{L}^\perp$는 $\mathbb{R}^n$의 부분 공간이며, $\mathcal{L} \cap \mathcal{L}^\perp = \{0\}$. (교집합이 없음)
- $\mathbb{R}^n$에 있는 x는 아래와 같이 표현 가능

  → $x = x_{\mathcal{L}} + x_{\mathcal{L}^\perp}$

  → $x_{\mathcal{L}} \in \mathcal{L}$ 이며, x를 $\mathcal{L}$로 정사영 시킨 벡터



$\|x - x_{\mathcal{L}}\|_2 \le \|x - y\|_2 \quad \forall y \in \mathcal{L}$

* Idempotent matrix

Ⓐ A → C(A)로 봄.
C(A)·x. → x를 A로 span 한 것.

- $A^2 = A$ 이라면 A는 Idempotent matrix라고 함.

  ① Ax는 x를 $C(A)$로 정사영 시킨 것. ($\forall x \in \mathbb{R}^n$) → Q. 근데 왜 C(A) 일까?

  ② $N(A) \perp C(A)$ 이며 $x^T y = 0 \quad \forall x \in N(A), \forall y \in C(A)$ → Q. N(A) vs C(A)?

  Ⓐ · N(A)는 A의 영공간.
  $Ax = 0$을 만족하는 벡터들

  ③ A는 대칭 행렬 & 역행렬

- 위 조건 만족하면 A는 C(A)로의 정사영 행렬이라 함.

· C(A)는 A의 행공간으로,
기저 역할을 함.
· x^T라고 하면 사실상 C(A)들의 조합이므로
영공간 A와 행공간 x의 내적은
음..

* Symmetric matrix

① $A^2 = A$

② A의 고유값은 0 또는 1

③ $rank(A) + rank(I_n - A) = n$

- 위 세가지 특징으로 아래 설명됨

  ① $rank(A) = tr(A)$ Q. rank(A), tr(A) 용어정리.

  ② $I_n - A$: $N(A)$로의 정사영

* Orthogonal projection

① A+B가 정사영인 경우.

- $\mathcal{C}(A+B) = \mathcal{C}(A) + \mathcal{C}(B)$
- $N(A+B) = N(A) \cap N(B)$
- $\boxed{A+B : \text{orthogonal projection} \iff \mathcal{C}(A) \perp \mathcal{C}(B)}$

(⇒ 증명) orthogonal 하므로.

$(A+B)^2 = A+B$, 이는 $AB+BA=0$ 임을 의미

$AB+BA=0$

$A^2B + ABA = AB + ABA = AB(I+A) = 0$ AB=0 이거나 I+A=0

$x \in \mathcal{C}(A)$ 이면 $Bx=0$ 임 ∨ 여기는 왜 증명?

x를 곱해서 $(I+A)ABx = 0 = AB(I+A)$

$\|Bx\|^2 = x^T B (I+A)^2 Bx = x^T B (I+3A) Bx = 0 \quad \therefore Bx = 0$

(⇐ 증명) $\mathcal{C}(A) \perp \mathcal{C}(B)$ 하므로

$(A+B)^2 = A^2 + B^2 + \cancel{AB} + \cancel{BA} = A+B$

∴ A+B는 orthogonal projection.

A, B : symetric → A+B : symetric

② AB가 orthogonal projection

- $\mathcal{C}(AB) = \mathcal{C}(A) \cap \mathcal{C}(B)$
- $N(AB) = N(A) + N(B)$
- [AB : orthogonal projection ⟺ AB=BA 증명.]

(⇒ 증명)

$(AB)^T = AB = B^T A^T \overset{\text{symetric}}{=} BA$

(⇐ 증명)

A가 대칭 & idem

$(AB)^2 = ABAB = AABB = AB$

③ T.F.A.E

- $\mathcal{C}(A) \subset \mathcal{C}(B)$
- $AB = A$
- $BA = A$
- $B-A$는 orthogonal projection

＊ 추가 내용.

- $X=(X_1, X_2, \ldots, X_p)$의 $n\times p$ 행렬이고, $H = X(X^TX)^{-1}X^T$ 이면

  H는 $\mathcal{C}(X)$로의 orthogonal projection 임.

  ① $H^2 = H$

  ② $\mathcal{C}(H) \perp N(H)$

  ③ $\mathcal{C}(H) = \mathcal{C}(X)$

- 만약 $n\times p$ 행렬인 Y가 $C(Y) = C(X)$ 일 때, $X(X^TX)^{-1}X^T = Y(Y^TY)^{-1}Y^T$

＊ 선형 회귀 – 벡터

- $x_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$ 일 때 $(x_1, y_1) \cdots (x_n, y_n)$이라 가정.
- 선형 회귀의 목적은 $\beta_0 + \sum_j \beta_j x_{ij}$ 로 y_i에 근사시키는 것.
- Least Square Estimator (LSE)는 $\underset{\beta_0, \beta_1, \ldots, \beta_p}{\text{minimize}} \sum_i \left(y_i - \beta_0 - \sum_j \beta_j x_{ij}\right)^2$

＊ 선형회귀 – 행렬

- $y = (y_1, \ldots, y_n)^T$, $y_i \in \mathbb{R}^n$ (n차원 벡터)
- X는 i번째 열이 $(1, x_i^T)$ 인 $n\times(p+1)$ 행렬 $\begin{bmatrix} \vdots & x_1^T & x_2^T & \cdots & x_n^T \end{bmatrix}$
- 행렬 식 일때는 LSE $\hat{\beta} = \underset{\beta \in \mathbb{R}^{p+1}}{\text{minimize}} \|y - X\beta\|_2^2 = (y-X\beta)^T(y-X\beta) = R(\beta)$
- 만약 $\text{rank}(X) = p+1$, LSE는 $\hat{\beta} = (X^TX)^{-1}X^Ty$.
- $\hat{y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty = Hy$ ($\hat{y}$는 y가 $\mathcal{C}(X)$로의 orthogonal projection).

· 만약 입력 변수가 $1^T x_j = \sum_i x_{ij} = 0$? $\forall j$, $1=(1,\dots,1)^T$ & $x_j=(x_{1j},\dots,x_{nj})^T$

· 이 경우에는 $\hat\beta_0 = \sum_i \frac{y_i}{n}$ → $(y-\hat\beta_0 1) \perp y$

· 만약, $x_1, x_2, \dots, x_p$가 orthogonal 하면, $\hat\beta_j = \frac{x_j^T y}{\|x_j\|_2^2}$ → $\frac{x\,x^T}{\|x\|_2^2}$: X로의 정사영

* 다중 회귀 by orthogonalization

· 선형회귀에서, 만약 p=1 이면, $\hat\beta_1 = \frac{(x_1-\bar x_1 1)}{\|x_1-\bar x_1 1\|_2^2} Y$, $\bar x_j = \sum_i \frac{x_{ij}}{n}$

nx2의 행렬

↳ Ax 가 x를 A로 내적(정사영)하는건 알겠음.

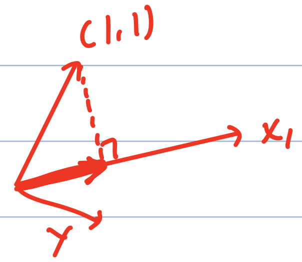
· 위 식은 아래처럼 적용가능.

↳ 근데 $x_1 - \bar x_1 1 = x_1 - \sum \frac{x_{i1}}{n}$ 은 뭐지? Q.

① 잔차인 $z = x_1 - \bar x_1 1$ 을 만들기 위해 x_1을 1로 회귀

② $\hat\beta_1$ 계수를 얻기 위해 y를 잔차 z로 회귀

· 이러한 접근은 p≥1 일 때 일반화됨.



* Gram-Schmidt orthogonalization

· 연속적인 직교화에 의한 회귀

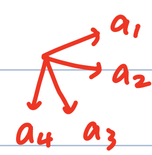
· $z_0 = x_0 = 1$

· for j=1,…,p 과정에서 x_j를 z_0부터 z_{j-1}에 회귀

· $\hat\gamma_{\ell j} = \frac{z_\ell^T x_j}{\|z_\ell\|_2^2}$

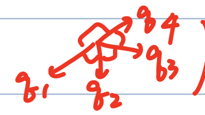
*그람-슈미트 직교화

$a_1 \quad a_2 \quad a_3 \quad a_4$

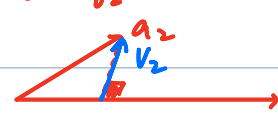


4차원

→ orthogonal basis를 찾기



다 크기가 1이고 서로서로 직교



$q_1 = \frac{v_1}{\|v_1\|_2}$ (normalize)

* QR decomposition

· $\Gamma = (\hat\gamma_{\ell j})$ 이고, $Z=(z_0,z_1,\dots,z_p)$, $D=\mathrm{diag}(\|z_j\|_2)$

이면, $X = Z\Gamma = Z\cdot D^{-1} D \cdot \Gamma \overset{\text{ref}}{=} QR$.

· $Q^TQ = I_{p+1}$ Q. 왜 p+1?

$v_1 = a_1$,

$v_2 = a_2 - \underbrace{\left(a_2^T \cdot \frac{v_1}{\|v_1\|_2}\right)}_{\text{정사영 크기}} \cdot \underbrace{\frac{v_1}{\|v_1\|}}_{\text{방향 벡터}}$

$= a_2 - \frac{a_2^T \cdot v_1}{\|v_1\|_2^2} v_1$

- Sampling distribution of $\hat{\beta}$ and $\hat{\sigma}$
  - X가 non-random한 선형 회귀에서는 아래 식을 가정함.

    $\varepsilon_i = y_i - x_i^T\beta \overset{iid}{\sim} N(0, \sigma^2)$

    ↳ 이 말은 즉슨, $y_i = x_i^T\beta + \varepsilon_i$ 이며 $\varepsilon_i \sim N(0, \sigma^2)$ 임.
  - $\text{rank}(X) = p+1$ 이고, $n > p+1$ 이라 가정함.
  - 모수 추정 결과, 다음과 같이 나옴
    - $\hat{\beta} = (X^TX)^{-1}X^Ty$

      $\hat{\sigma}^2 = \dfrac{\|y-\hat{y}\|_2^2}{n-p-1}$
  - $Z \sim N(0_n, I_n)$ 이고 $A \in \mathbb{R}^{n\times n}$ 이 직교 행렬 일 때 아래 식을 만족함.

    $Z^TAZ \sim \chi_k^2$.
    - (증명)
    - $Z \sim N(0_n, I_n)$
    - $A = VV^T$ (spectral decomp.) $= v_1v_1^T + v_2v_2^T + \cdots + v_kv_k^T$.
    - $Z^TAZ = (v_1^TZ)^2 + (v_2^TZ)^2 + \cdots + (v_kZ)^2 \sim \chi_k^2$

$A = VDV^T$

Z^TAZ

$= Z^TVDV^TZ$

      - $E[v_j^TZ] = 0$, $Var[v_j^TZ] = v_j^T Var(Z) v_j = 1$

$V^TZ \sim N(0, V^T Var(Z) V)$ $Cov(v_j^TZ, v_i^TZ) = v_j^T Cov(Z, Z) v_i^T = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

$= N(0_n, I_n)$

$\therefore Z^TAZ = \tilde{Z}^TD\tilde{Z}$ $(\tilde{Z} = V^TZ) \sim \chi_k^2$

  - $y_i = x_i^T\beta + \varepsilon_i$, $Y = X\beta + \Sigma$, $\Sigma \sim N(0, \sigma^2 I_n)$
  - $\hat{\beta} = (X^TX)^{-1}X^Ty = (X^TX)^{-1}X^T(X\beta + \Sigma)$

    $= \beta + (X^TX)^{-1} \cdot X^T \cdot \Sigma$ → 여기서 $\Sigma \sim N(0, (X^TX)^{-1}X^T \cdot \sigma^2 \cdot X \cdot (X^TX)^{-1})$

    $= N(0, \sigma^2 \cdot (X^TX)^{-1})$
  - $\hat{\sigma}^2 =$