if $x, y \in S$, then $\alpha x + \beta y \in S$ for any α, β.

ex) recommender system

$X = UV$

$$X = \left[\quad \right]_{p}^{n} \quad U = \left[\quad \right]_{p}^{r} \quad V = \left[\quad \right]_{r}^{n}$$

U: ↳ r representative taste profiles.

V: ↳ r weights for each user.

The span of columns of U is a subspace.
This means that all columns of X lie in that subspace

↳ X의 column은 U의 column에 가중치를 곱해 더한 값이므로
X의 column은 U의 column이 span한 subspace 안에 있다.

↳ $x, y \in S$, then $\alpha x + \beta y \in S$ for any α, β.

How to represent a subspace?

- Represent S as the span of a set of vectors.
- Represent S as the span of a set of linearly independent vector (basis)
- Represent S as the span of a set of orthonormal vectors (orthonomal basis)

Let U is an orthonormal matrix

$$U^T U = C \Rightarrow C_{ij} = \langle u_i, u_j \rangle$$

$$\langle u_i, u_j \rangle = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

$$\therefore C = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} = I$$

ex) $\|Uv\|_2^2$

$$= (Uv)^T (Uv)$$

$$= v^T U^T U v = v^T v$$

$$\therefore \|Uv\|_2^2 = \|v\|_2^2$$

$\dim(S)$ = 기저벡터의 갯수

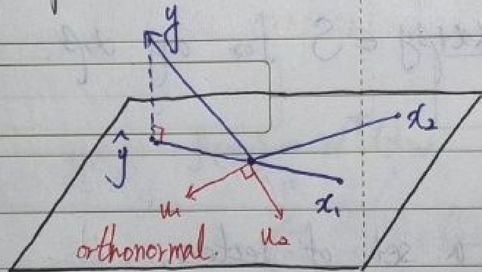
if $S = \text{span}(\text{col}(X))$, then $\dim(S) = \text{rank}(X)$

$\hat{y}$ = Projection of y onto set S $= P_S y = \underset{x \in S}{\text{argmin}} \|y - x\|_2^2$

$\hat{y} = X\hat{w}$, $\hat{w} = (X^T X)^{-1} X^T y$ (Least squares)

$\Rightarrow \hat{y} = X(X^T X)^{-1} X^T y = P_X y$ (projection matrix)

$\text{span}(\text{cols}(U)) = \text{span}(\text{cols}(X))$



$w_1 x_1 + w_2 x_2$

$= \tilde{w}_1 u_1 + \tilde{w}_2 u_2$

↳ 같은 값을 다른 식으로 표현할 수 있다.

↳ 계산이 쉬운 정규직교 벡터들에 대한 식으로 바꾼다.

$\hat{y} = X\hat{w} = U\tilde{w}$ $\quad \tilde{w} = \underset{w}{\text{argmin}} \|y - Uw\|_2^2 = (U^T U)^{-1} U^T y$

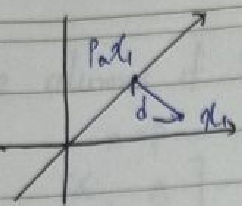
$\therefore \hat{y} = U(\underline{U^T U})^{-1} U^T y$ $\quad \Rightarrow \boxed{\hat{y} = UU^T y}$

↳ U is orthobasis

↳ $U^T U = I$

gongs!n

a (subspace)



$$d_i^2 = \| x_i - P_a x_i \|_2^2$$

Properties of projection Matrix

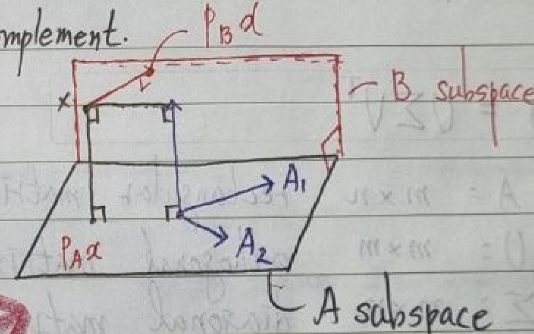
- $P_A = P_A^2 = P_A P_A$
- $P_A = P_A^T = P_A^T P_A$
- if $A = a$ (single column Vector), then $P_a = a(a^Ta)^{-1}a^T = \frac{aa^T}{a^Ta}$ (Scaler)

Orthogonal Complement (직교여공간)

주어진 부분공간과 수직인 벡터들의 공간

Let B be a basis for orthogonal Complement.

$$\boxed{A^T B = 0}$$



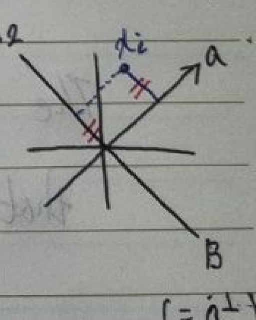
For any vector $x \in \mathbb{R}$, it can be written as

$$x = P_A x + P_B x = (P_A + P_B)x$$

$\hookrightarrow$ $Ix = (P_A + P_B)x \Rightarrow I = P_A + P_B \Rightarrow P_B = I - P_A$

$$d_i^2 = \| x_i - P_a x_i \|_2^2 = \| x_i - \frac{aa^T}{a^Ta} x_i \|_2^2 = \| (I - \frac{aa^T}{a^Ta}) x_i \|_2^2$$

$\hookrightarrow$ $P_B x_i$



$$= x_i^T (I - \frac{aa^T}{a^Ta})^T (I - \frac{aa^T}{a^Ta}) x_i$$

$$= x_i^T (I - \frac{aa^T}{a^Ta}) x_i \quad (P_A^T P_A = P_A)$$

$$= x_i^T x_i - \frac{x_i^T aa^T x_i}{a^Ta}$$

OXFORD

a에 관해 최소화

$$\sum_{i=1}^{p} d_i^2 = \sum_{i=1}^{p} x_i^T x_i - \frac{x_i^T a a^T x_i}{a^T a}$$

$$\hat{a} = \underset{a}{\operatorname{argmin}} \sum_{n=1}^{p} d_i^2 = \underset{a}{\operatorname{argmin}} \sum_{n=1}^{p} - \frac{x_i^T a a^T x_i}{a^T a}$$

$$= \underset{a}{\operatorname{argmax}} \sum_{i=1}^{p} \frac{x_i^T a a^T x_i}{a^T a} = \underset{a}{\operatorname{argmax}} \sum_{i=1}^{p} \frac{a^T x_i x_i^T a}{a^T a}$$

($x_i^T a, a^T x_i$ are scala)

$$\boxed{\hat{a} = \underset{a}{\operatorname{argmax}} \frac{a^T X X^T a}{a^T a}}$$ 1st left singular Vector

$\hookrightarrow \sigma_i^2$ is the squared 1st singular value of X

$$\boxed{A = U \Sigma V^T}$$

A : $m \times n$ rectangular matrix
U : $m \times m$ orthogonal matrix (left singular vertor)
Σ : $m \times n$ diagonal matrix
V : $n \times n$ orthogonal matrix

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

U give an ortho~~gonal~~ normal. basis for A

The 1st column is the best 1D line that fits to all data.

The columns of V are called the right singular vector

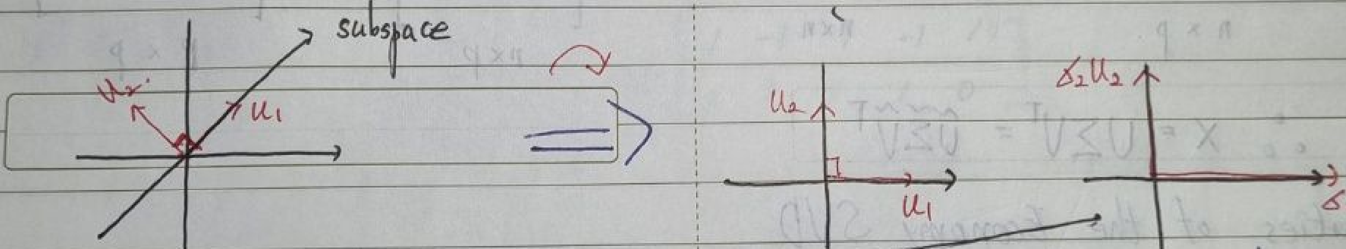
$$\begin{array}{ccc} n=p & n>p & n<p \\ \Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ & & \ddots & \\ 0 & & & \sigma_n \end{bmatrix} & \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \\ \hdashline & 0 & \end{bmatrix} & \left[\begin{array}{ccc:c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_n & \end{array}\right] \end{array}$$

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$$

$u_1 = \underset{a}{\operatorname{argmax}} \dfrac{a^T X X^T a}{a^T a}$ / $\tilde{x}_i^{(1)} = x_i - P_{u_1} x_i$ (project x_i onto u_i)

$\tilde{x}_i^{(2)} = x_i - P_{u_1} x_1 - P_{u_2} x_1$

$\vdots$



Any point $x = \sigma_1 u_1 v_1 + \sigma_2 u_2 v_2$

singular value σ_n indicate how spread out points are in Subspace

Using the u_i's rotate
σ_is rescale
v's 위치

Σ의 개수 = rank(X)

if X has SVD, we can write X as sum of rank-1 matrices.

$$X = \sum_{i=1}^{r=\min(n,p)} \sigma_i u_i v_i^T$$

The economy SVD

Assume $X \in R^{n\times p}$ has $\text{rank}(X) = r \lll \min(n,p)$

$$\underset{n\times p}{X = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}} \quad \underset{n\times n}{U = \begin{bmatrix} \boxed{\tilde{U}} & \end{bmatrix}} \quad \underset{n\times p}{\Sigma = \begin{bmatrix} \boxed{\begin{matrix} \sigma_1 & 0 \\ 0 & \sigma_k \end{matrix}} & 0 \\ 0 & 0 \end{bmatrix}} \quad \underset{p\times p}{V^T = \begin{bmatrix} \boxed{\tilde{V}^T} \\ \\ \end{bmatrix}}$$

$$\boxed{\therefore\ X = U\Sigma V^T = \tilde{U}\tilde{\Sigma}\tilde{V}^T}$$

Properties of the Economy SVD

$\sigma_1 \ge \sigma_2 \cdots \ge \sigma_r > 0$

$\tilde{U}^T U = I,\ \tilde{U}\tilde{U}^T \ne I,\ \tilde{V}^T\tilde{V} = I,\ \tilde{V}\tilde{V}^T \ne I$

차원축소 : SVD를 이용하여 고차원의 데이터를 유용한 저차원의 데이터로 변환

not SVD / not orthogonal

Ex)

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

normalize

$$= \begin{bmatrix} 1/\sqrt{5} & 1/2 \\ -1/\sqrt{5} & 1/2 \\ 1/\sqrt{5} & 1/2 \\ -1/\sqrt{5} & 1/2 \\ 1/\sqrt{5} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{5}\sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

As the rank of X is 2, all of the columns of X lie in a 2D subspace.

$\tilde{\Sigma}\tilde{U}^T$ (tells us where they lie in that S)

$$X^T = \tilde{V}\tilde{\Sigma}\tilde{U}^T = \underbrace{\tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & -1 \\ 0 & 1 \end{bmatrix}}_{\tilde{V}} \cdot \sqrt{2} \begin{bmatrix} 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

x_i = weighted sum of columns of $\tilde{V}$

z_i = reduced dimensional vectors = weights.

$X = U\Sigma V^T$ $\qquad \hat{w} = (X^TX)^{-1}X^Ty$

$$(X^TX)^{-1}X^T$$
$$= (V\Sigma^TU^TU\Sigma V^T)^{-1}V\Sigma^TU^T$$
$$= (V\Sigma^T\Sigma V^T)^{-1}V\Sigma^TU^T$$
$$= (V^T)^{-1}(V\Sigma^T\Sigma)^{-1}V\Sigma^TU^T$$
$$= (V^{-1})^{-1}(V\Sigma^T\Sigma)^{-1}V\Sigma^TU^T$$
$$= V(\Sigma^T\Sigma)^{-1}V^{-1}V\Sigma^TU^T$$
$$= V(\Sigma^T\Sigma)^{-1}\Sigma^TU^T$$

If A, B are square and invertible then $(AB)^{-1}$, $B^{-1}A^{-1}$

$V^TV = VV^T = I = V^{-1}V$

$\Rightarrow V^T = V^{-1}$

$(\Sigma^T\Sigma)^{-1}\Sigma^T \quad (n \geq p)$

$\Sigma\ (n \times p)$

$$\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \\ \hline & 0 & \end{bmatrix}$$

$\Sigma^T\Sigma\ (p \times p)$

$$\begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_p^2 \end{bmatrix}$$

$(\Sigma^T\Sigma)^{-1}$

$$\begin{bmatrix} 1/\sigma_1^2 & & & 0 \\ & 1/\sigma_2^2 & & \\ & & \ddots & \\ 0 & & & 1/\sigma_p^2 \end{bmatrix}$$

$(\Sigma^T\Sigma)^{-1}\Sigma^T\ (p \times n)$

$$\begin{bmatrix} 1/\sigma_1 & & & 0 & \vdots & \\ & 1/\sigma_2 & & & \vdots & 0 \\ & & \ddots & & \vdots & \\ 0 & & & 1/\sigma_p & \vdots & \end{bmatrix} \rightarrow \Sigma^+$$

pseudo-inverse

$\therefore\ (X^TX)^{-1}X^T = V\Sigma^+U^T$

$\hat{y} = X\hat{w} = U\Sigma V^T\hat{w}$

$\hat{w} = V\Sigma^+U^Ty$

$$\Rightarrow \quad y \approx \hat{y} = U\Sigma V^T w$$

$$\Downarrow$$

$$U^Ty \approx \Sigma V^T w$$

$$\Downarrow$$

$$\Sigma^+U^Ty \approx \Sigma^+\Sigma V^T w \quad (\Sigma^+\Sigma = I)$$

$$\boxed{V\Sigma^+U^Ty \approx \hat{w}}$$