기초 선형대수
(Spectral Property – 고유값/고유벡터/행렬의 인수분해)

윤 정 훈

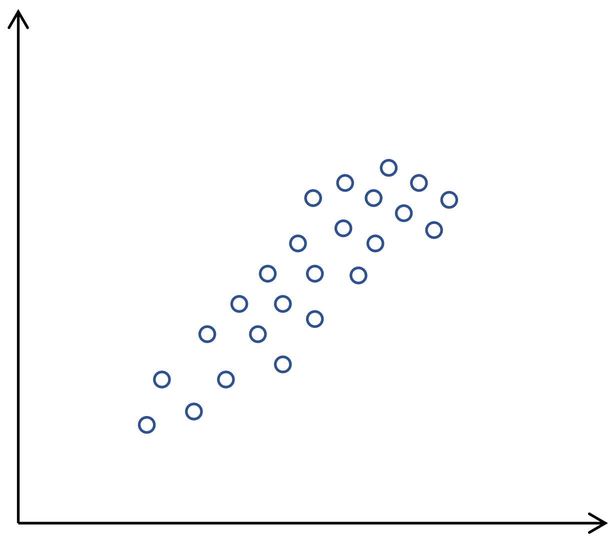
들어가며

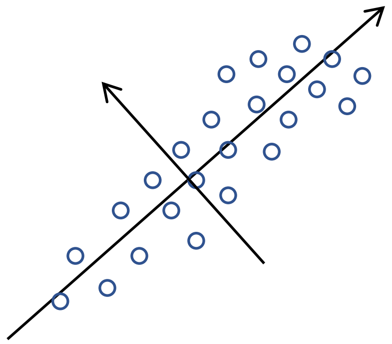
Question 1)

데이터 집합이 다음과 같다.

어떤 것이 더 나은 표현인가?

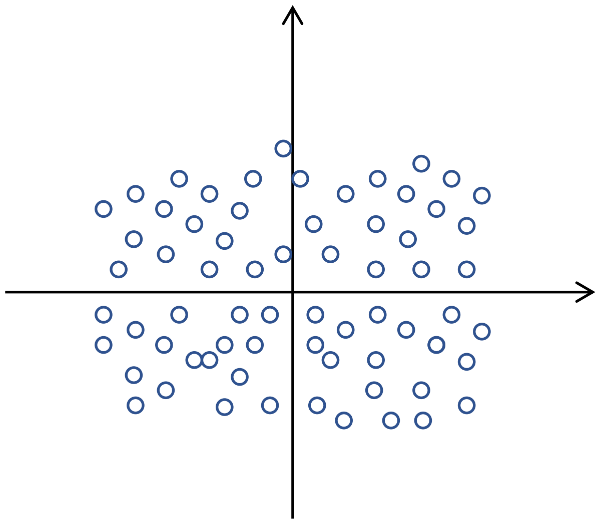
더 나은 표현이라는 것은 무슨 의미인가?

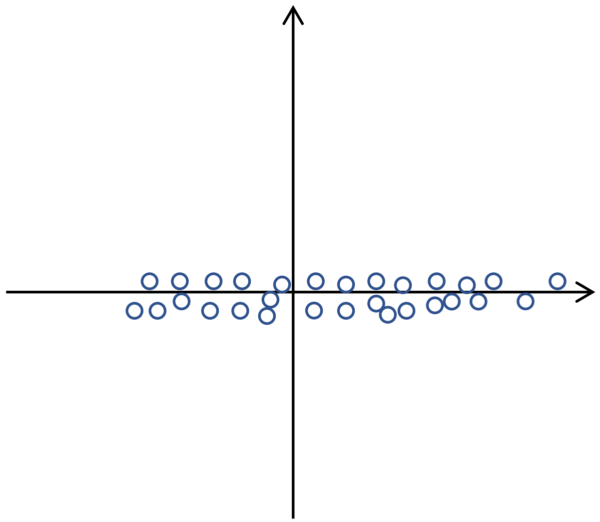




들어가며

Question 2) 데이터 집합을 표현할 때, x축 y축이 모두 필요한가?

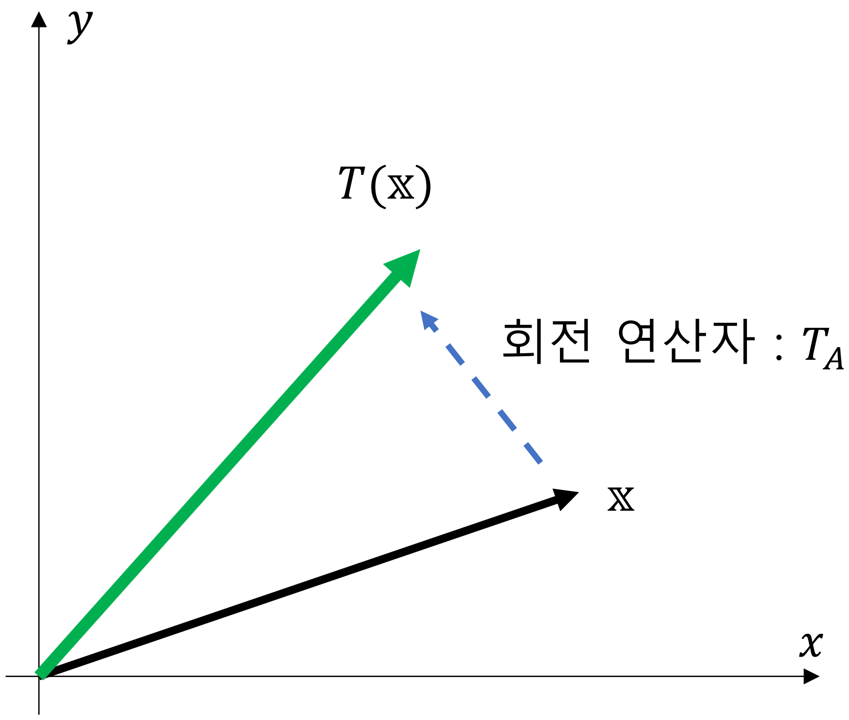




다양한 선형변환

- $R^n \longrightarrow R^m (R^n)$으로의 대표적인 선형변환
  - 회전변환
  - 반사변환
  - 사영변환

$R^2 \longrightarrow R^2$의 회전변환

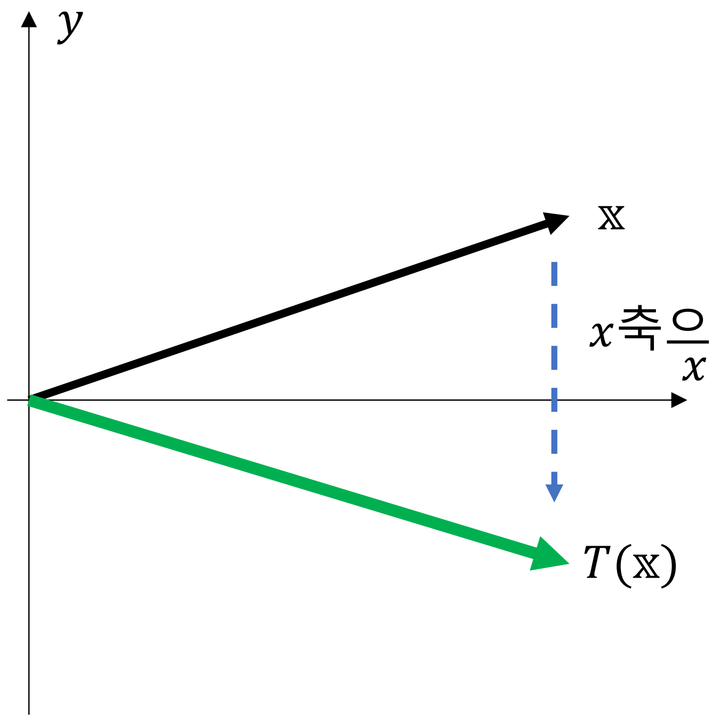


회전 연산자 : $T_A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

회전연산자의 특징

1. 벡터의 방향이 변한다.
2. 벡터의 길이는 변하지 않는다.

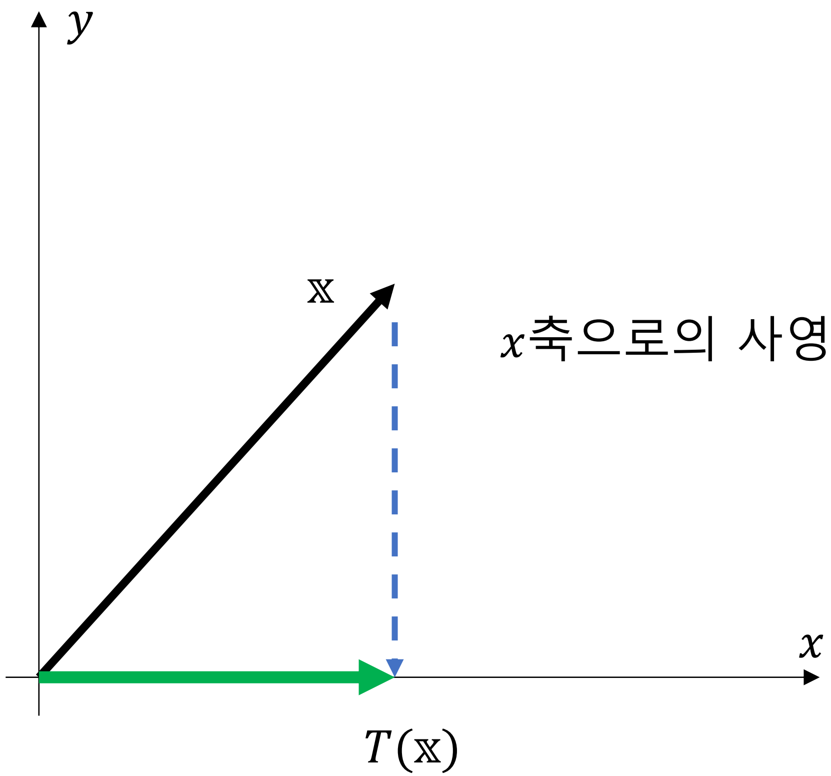
$R^2 \longrightarrow R^2$의 반사변환



x축으로의 반사 연산자 : $T_A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

반사연산자의 특징
1. 벡터의 방향이 변한다.
2. 벡터의 길이는 변하지 않는다.

$R^2 \longrightarrow R^2$의 사영변환



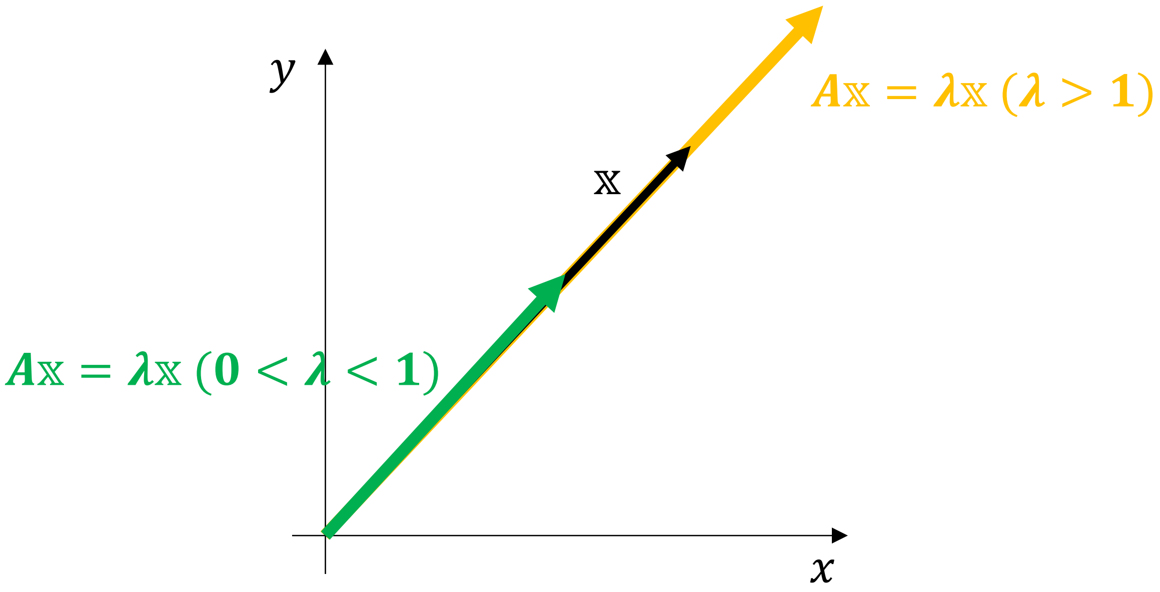
x축으로의 사영연산자 : $T_A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

회전연산자의 특징
1. 벡터의 방향이 변한다.
2. 벡터의 길이도 변하다.

그러면 **선형변환의 결과**로 벡터의 방향은 바뀌지 않고, 벡터의 길이가 바뀌는 선형변환은 무엇이 있을까?

고유값과 고유벡터

- 정방행렬 A 에 대하여, 스칼라(scala)인 λ 와 영이 아닌 벡터 $\mathbb{v}$ 에 대해 $\boldsymbol{A}\mathbb{v} = \boldsymbol{\lambda}\mathbb{v}$ 를 만족하는 경우, λ 는 A 의 고유값(eigenvalue) , $\mathbb{v}$ 는 대응하는 고유벡터(eigenvector) 라고 한다.
- λ 가 행렬 A 의 고유값이면, 대응하는 고유벡터는 무수히 많다.
- 집합 $\{\mathbb{v}: A\mathbb{v} = \lambda\mathbb{v}\}$ 는 벡터공간이며 고유값 λ 에 대응하는 고유공간(eigenspace)이라 한다.



고유값과 고유벡터 예제

- $A = \begin{bmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{bmatrix}$의 고유값 중 하나는 $\lambda = 5$이고, 이에 대응하는 고유벡터 $\mathbb{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$이다.

- $A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & 1 & 2 \end{bmatrix}$의 고유값 중 하나는 $\lambda = 0$이고, 이에 대응하는 고유벡터 $\mathbb{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$이다.

- A 를 대각행렬이라고 하자. $A = diag(\lambda_1, \dots, \lambda_n)$ 의 고유벡터와 고유값은? 표준 기저 벡터 $\mathbb{e}_1, \dots, \mathbb{e}_n$ 는 고유벡터이고, 대각원소인 $\lambda_1, \dots, \lambda_n$ 은 고유값이다.

고유값과 고유벡터의 계산

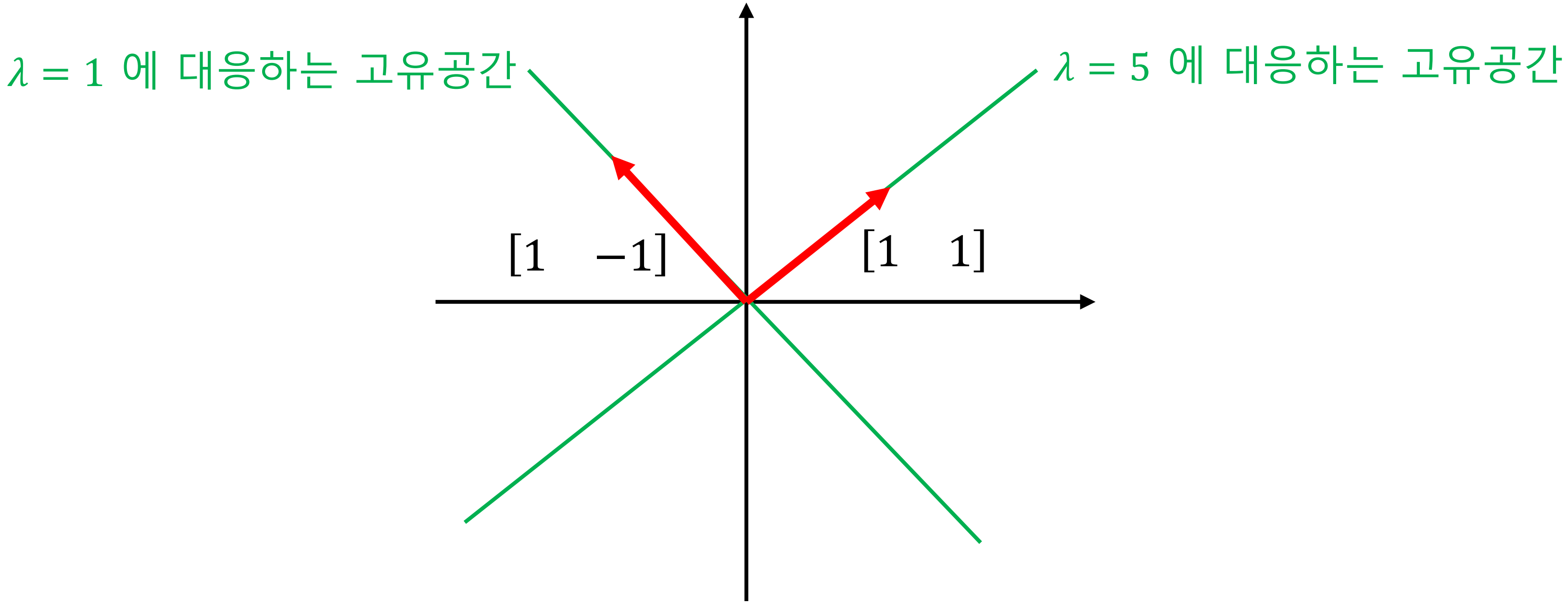
행렬 A 의 고유값을 λ , 대응하는 고유벡터를 $\mathbb{v}$ 라고 하자.

- $A\mathbb{v} = \lambda\mathbb{v} \rightarrow (A - \lambda I_n)\mathbb{v} = \mathbb{O}$ 즉, 특성방정식 $\det(A - \lambda I_n) = 0$ 을 만족시키는 λ 를 구하면 된다.
  - 2×2 의 경우는 $\det(A - \lambda I_2) = \lambda^2 - tr(A)\lambda + \det(A) = 0$ 을 만족시켜주는 λ를 구하면 된다.

- 행렬 $A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$ 의 고유값과 그에 대응하는 고유벡터?

고유벡터 예제

- 행렬 $A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$ 의 고유값과, 그에 대응하는 고유벡터를 구하여라. 그리고 고유공간을 그래프로 그려라.
  - $tr(A) = 6$, $\det(A) = 5$ 이므로 $\lambda^2 - 6\lambda + 5 = (\lambda - 1)(\lambda - 5) = 0$ 을 만족시키는 λ의 값은 1, 5 이다.



유사성

- 역행렬이 존재하는 행렬을 가역행렬이라고 한다.

- 가역행렬 S 에 대해 $\boldsymbol{S^{-1}AS = B}$가 만족되면 두 정방행렬 A와 B는 "**유사** 혹은 **닮은**(**similar**) **행렬**"이라고 한다.

- 유사행렬(similar matrix)들은 동일한 고유값을 가진다.

대각화 가능성

- 만약 어떤 **정방행렬 $\boldsymbol{A}$ 가 대각행렬 $\boldsymbol{D}$ 와 유사행렬**이면, 즉 대각행렬 D 에 대해 $\boldsymbol{S^{-1}AS = D}$를 만족하는 가역행렬 S가 있으면, A는 "**대각화 가능하다(diagonalizable)**" 라고 한다.

- $n\times n$행렬 A가 대각화 가능 $\Leftrightarrow$ A가 n개의 일차독립인 고유벡터를 갖는 것

- n개의 일차독립인 고유벡터를 갖는 $n\times n$행렬 A를 대각화하는 법
  - A 의 n 개의 일차독립인 고유벡터 $\mathbb{s}_1, \dots, \mathbb{s}_\mathrm{n}$ 을 구한다.
  - 행렬 $S = [\mathbb{s}_1 \quad \cdots \quad \mathbb{s}_n]$ 을 구성한다.
  - 행렬 $S^{-1}AS$ 는 $\mathbb{s}_1, \dots, \mathbb{s}_n$ 에 각각 대응하는 고유값을 대각성분으로 하는 대각행렬이 될 것이다.

대각화 가능성 예제

행렬 $A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$ 은 고유값 $\lambda = 1$, $\lambda = 2$ 를 갖는다.

$\lambda = 1$ 에 대응하는 고유공간의 기저벡터는 $\mathbb{s}_1 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$

$\lambda = 2$ 에 대응하는 고유공간의 기저벡터는 $\mathbb{s}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbb{s}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ 이다.

3개의 기저벡터는 일차독립이므로 A 는 대각화 가능하고 $S = \begin{bmatrix} -2 & -1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

$S^{-1}AS = \begin{bmatrix} -1 & 0 & -1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -2 & -1 & 0 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = D$ 로 대각화 된다.

대각화 가능과 고유값 사이의 관계

- 먼저, 대각행렬 $D = diag(\lambda_1, \dots, \lambda_n)$ 의 행렬의 고유값은 $\lambda_1, \dots, \lambda_n$ 이다. **행렬 $\boldsymbol{A}$ 와 $\boldsymbol{D}$ 가 유사행렬**이라고 하자. 즉, 가역행렬 S 에 대하여, $\boldsymbol{S^{-1}AS = D}$ 라고 하자. 그리고 D 가 대각행렬이므로, **$\boldsymbol{A}$ 는 대각화 가능**하다. A 의 고유값은 D 의 고유값, 즉 D 의 대각원소들이다. 그리고 S 의 열들은 일차 독립이며 A 의 고유벡터들이 된다.

직교대각화 가능

- 정사각행렬 S에 대하여, $\boldsymbol{S}^{T} = \boldsymbol{S}^{-1}$인 경우, A를 **직교행렬**이라고 한다.

- 정사각행렬 A에 대해서, $\boldsymbol{D} = \boldsymbol{S}^{T}\boldsymbol{A}\boldsymbol{S}$를 만족하는 대각행렬 D와, 직교행렬 S가 존재하면 A는 직교대각화 가능(orthogonally diagonalizable) 하다고 한다.

- 정사각행렬 A 가 직교대각화 가능 $\Leftrightarrow$ 정사각행렬 A 가 대칭행렬

행렬의 인수분해

정수는 인수분해된다. 예를들어 $12 = 3 \times 4$처럼, 12는 3과 4로 인수분해된다.

마찬가지로 행렬도 $A = BC$처럼 인수분해 될 수 있다.

예를들어, $\begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ 이렇게 인수분해 될 수 있다.

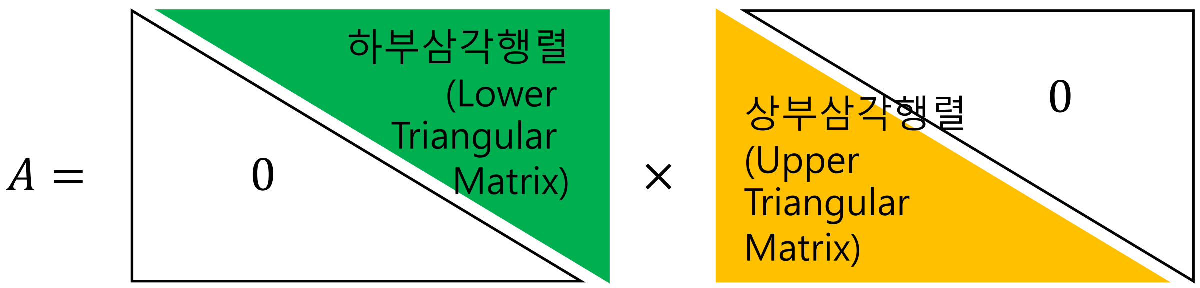
행렬 A를 인수분해하여, 행렬 A와 가장 비슷하고(similar), 크기가 작은 행렬을 A대신 사용하려고 한다.

여러가지 행렬의 인수분해

- LU 분해 (LU Decomposition)
- 스펙트럼 분해 (Spectrum Decomposition) or 고유값 분해
- 특이값 분해 (Singular Value Decomposition)

LU 분해

- LU분해는 행렬을 아래의 두 행렬로 인수분해한다.
  - 하부삼각행렬(행렬의 대각선의 윗부분이 모두 0)
  - 상부삼각행렬(행렬의 대각선의 아랫부분이 모두 0)



- LU분해를 통하여 Linear Systems 문제를 간단하게 해결할 수 있다.

LU 분해 예제

- $\begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$ 를 쉽게 풀어보자!

하부삼각행렬 상부삼각행렬

$\begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix}\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ 처럼 삼각행렬의 곱으로 이루어진다.

$\begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix}\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$ 처럼 풀고자하는 식을 다시 쓸수 있다.

먼저, $\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ 으로 정의하자.

LU 분해 예제

그러면 우리가 풀고자 하는 식은, $\begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$ 으로 다시 쓸 수 있다.

즉, $2y_1 = 2$, $-3y_1 + y_2 = 2$, $4y_1 - 3y_2 + 7y_3 = 3$ 을 풀면 된다. 즉 $y_1 = 1$, $y_2 = 5$, $y_3 = 2$ 가 된다.

식을 다시 써 보면, $\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix}$ 가 된다.

즉, $x_1 + 3x_2 + x_3 = 1$, $x_2 + 3x_3 = 5$, $x_3 = 2$ 을 풀면 된다. 즉 $\boldsymbol{x_1 = 2}$ **,** $\boldsymbol{x_2 = -1}$ **,** $\boldsymbol{x_3 = 2}$ 가 된다.

이 예제는 행렬 A를 하부삼각행렬, 상부삼각행렬로 인수분해하면, 선형계 $A\mathbb{x} = \mathbb{b}$ 는 풀이과정이 쉬워진다는 것을 명확히 보여주고 있다.

LU 분해

- 정사각행렬 A가 하부삼각행렬 L과 상부삼각행렬 U의 곱 $A = LU$ 로 되는 인수분해를 LU 분해 또는 LU 인수분해(LU factorization) 라고 부른다.

- 일반적으로 모든 정사각행렬 A가 LU 분해를 갖지 않으며, LU 분해가 존재하여도 LU 분해가 유일하지도 않다.

스펙트럼 분해

- 행렬 A가 $S = [\mathbb{s}_1 \quad \cdots \quad \mathbb{s}_n]$로 직교대각화되는 대칭행렬이고, $\lambda_1, \dots, \lambda_n$은 각각 $\mathbb{s}_1, \dots, \mathbb{s}_n$에 대응하는 A의 고유값이라고 하면, 행렬 $D = S^T A S$는 대각성분이 A의 고유값인 대각행렬이 된다.

- $A = SDS^T = [\mathbb{s}_1 \quad \cdots \quad \mathbb{s}_n] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbb{s}_1^T \\ \vdots \\ \mathbb{s}_n^T \end{bmatrix} = \lambda_1 \mathbb{s}_1 \mathbb{s}^T + \cdots + \lambda_n \mathbb{s}_n \mathbb{s}_n^T$ 와 같이 표현 가능하다.

- 이 공식을 A 의 스펙트럼 분해(spectral decomposition) 또는 A의 고유값 분해(eigenvalue decomposition) 라고 한다.

특이값 분해

- SVD는 모든(any) 행렬에 대하여 적용가능한 분해(정수와 비교하면 인수분해) 기법이기 때문에, 행렬의 분해에 가장 널리 사용되는 기법이다.

- $n \times n$ 행렬 A 가 대칭행렬이 아니면, 고유값 분해는 존재하지 않는다. 일반적인 경우에 대한 행렬의 인수분해를 위해서 특이값 분해(Singular Value Decomposition)를 사용한다.

- 디지털화된 정보의 압축, 저장, 전송에 활용

특이값 분해

- 행렬 A가 계수 k인 $m{\times}n$행렬일 때, $\boldsymbol{A = U\Sigma V^T}$로 인수분해된다.

$$A = U\Sigma V^T = \begin{bmatrix} \mathbb{u}_1 & \cdots & \mathbb{u}_k & | & \mathbb{u}_{k+1} & \cdots & \mathbb{u}_m \end{bmatrix} \left[\begin{array}{ccc:c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O_{k\times(n-k)} \\ 0 & \cdots & \sigma_k & \\ \hdashline & O_{(m-k)\times k} & & O_{(m-k)\times(n-k)} \end{array}\right] \begin{bmatrix} \mathbb{v}_1^T \\ \vdots \\ \mathbb{v}_k^T \\ - \\ \mathbb{v}_{k+1}^T \\ \vdots \\ \mathbb{v}_n^T \end{bmatrix}$$

- U : $m{\times}m$, Σ : $m{\times}n$, V : $n{\times}n$
- $V = [\mathbb{v}_1 \quad \cdots \quad \mathbb{v}_n]$ 는 A^TA 를 직교대각화한다. 즉, **$\boldsymbol{V}$ 의 컬럼은 $\boldsymbol{A^TA}$ 의 고유벡터들**이다. $\mathbb{v}_i$는 orthonomal.
- $\lambda_1, \ldots, \lambda_k$ 가 V 의 열벡터에 대응하는 **$\boldsymbol{A^TA}$ 의 영이 아닌 고유값**일 때, Σ 의 영이 아닌 대각성분은 $\boldsymbol{\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_k = \sqrt{\lambda_k}}$ 이다. 이 대각성분들을 **특이값(singular value)**이라고 한다. k는 A의 계수이다.
- V 의 열벡터는 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$ 을 만족시키도록 배열되어 있다.
- $\{\mathbb{u}_1, \ldots, \mathbb{u}_k\}$ 는 col(A)의 정규직교기저이다. $\mathbb{u}_i$는 orthonormal.

$A^T A$, AA^T

- $A^T A$ (임의의 행렬 A)
  - $A^T A = V\Sigma^T U^T U \Sigma V^T = V(\Sigma^T \Sigma)V^T = VDV^T$
    - $D = \Sigma^T \Sigma$ 이기 때문에 D의 대각성분은 singular value의 제곱
  - $(A^T A)V = VDV^T V = VD$
  - 위의 식에 의하여 $A^T A$의 고유값은 V의 대각성분들이 되고, V의 columns은 $A^T A$의 고유벡터가 된다.
- AA^T (임의의 행렬 A)
  - $AA^T = U\Sigma V^T V \Sigma^T U^T = U\Sigma^T \Sigma U^T = UDU^T$
    - $D = \Sigma^T \Sigma$ 이기 때문에 D의 대각성분은 singular value의 제곱
  - $(AA^T)U = UDU^T U = UD$
  - 위의 식에 의하여 AA^T의 고유값은 U의 대각성분들이 되고, U의 columns은 AA^T의 고유벡터가 된다.
- $A^T A$의 eigenvector는 PCA에서의, A의 공분산 행렬의 주성분이다.

특이값 분해 예제

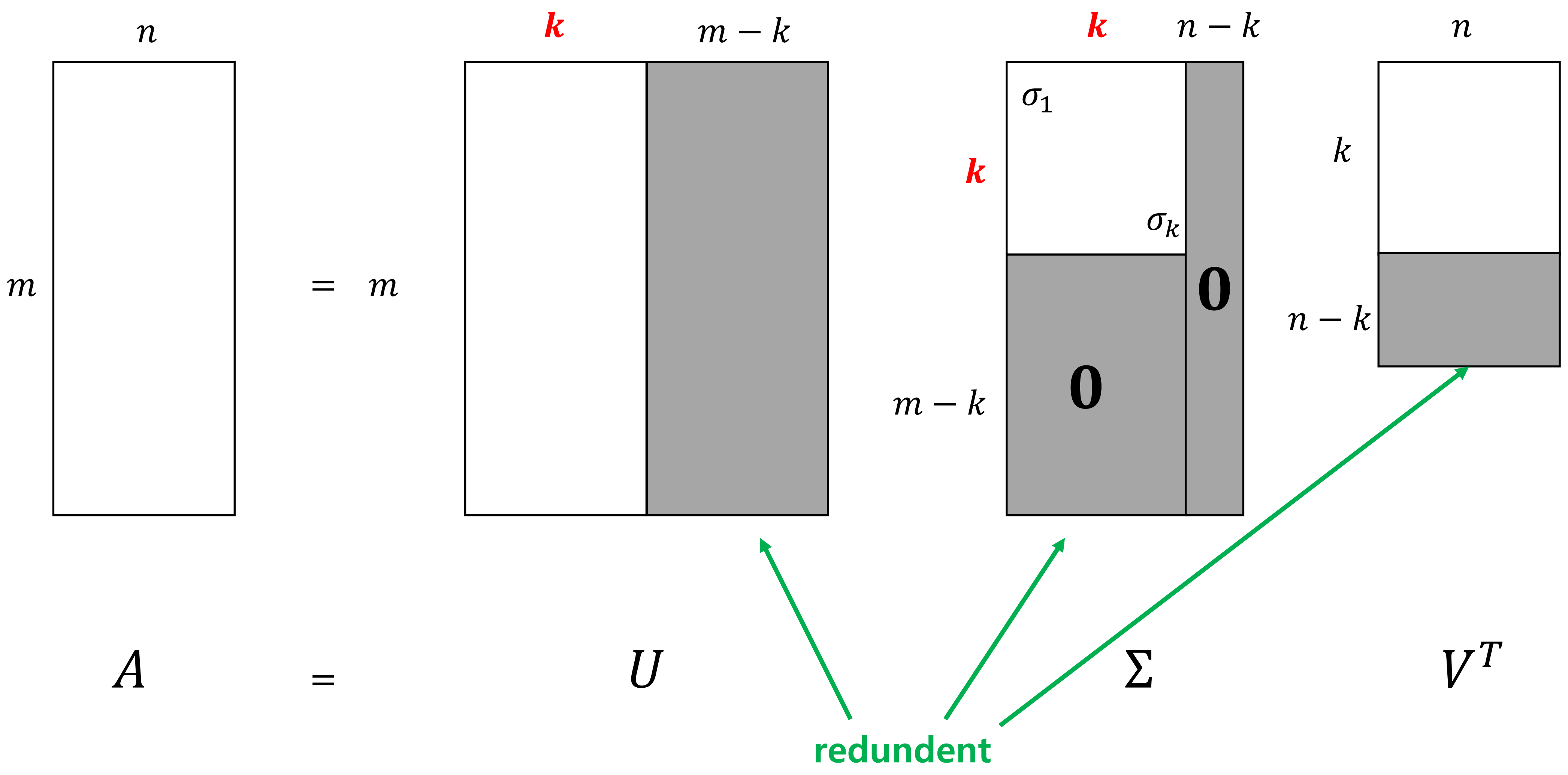
- 행렬 $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$ 의 특이값분해를 구하라.
  - $A^T A$ 의 고유값을 구한다. $\lambda_1 = 3$, $\lambda_2 = 1$
  - 특이값을 구한다. $\sigma_1 = \sqrt{\lambda_1} = \sqrt{3}$, $\sigma_2 = \sqrt{\lambda_2} = \sqrt{1}$
  - $A^T A$ 의 고유벡터를 구한다. $(\mathbb{v}_1, \dots, \mathbb{v}_n)$
  - $\frac{1}{\sigma_i} A\mathbb{v}_i = \mathbb{u}_i$를 구한다. $(\mathbb{u}_1, \dots, \mathbb{u}_m)$

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

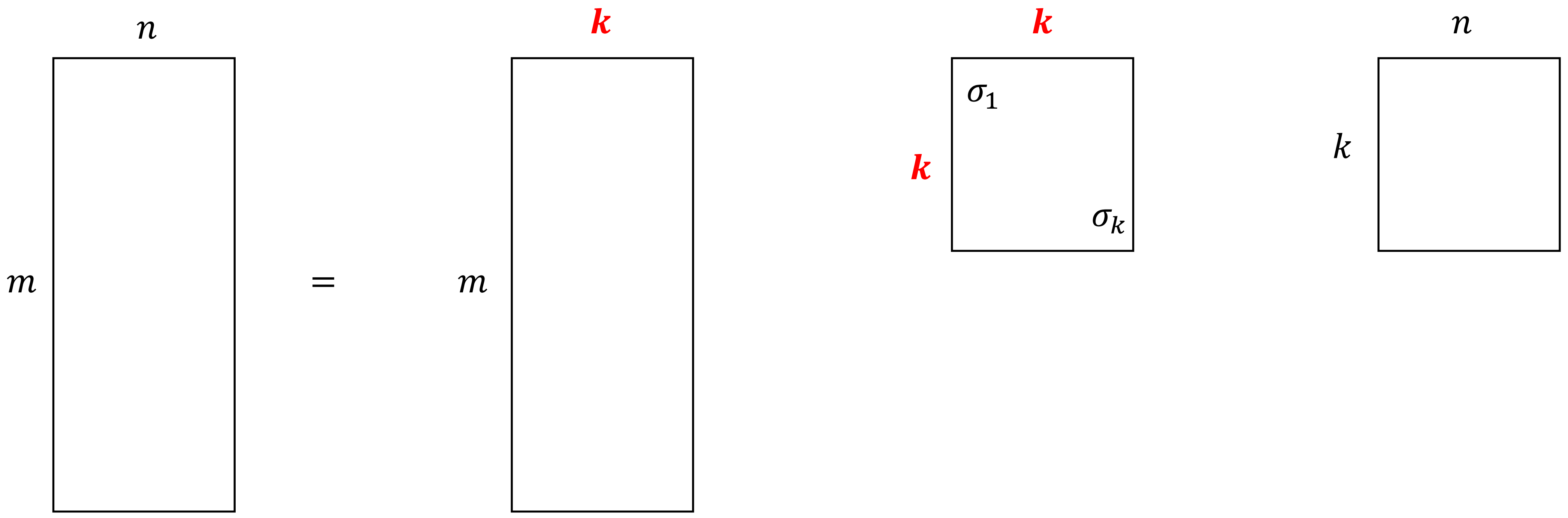
$$= U\Sigma V^T$$

$$= \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

축소된 SVD



축소된 SVD



$$A = U_1 \Sigma_1 V_1^T$$

축소된 SVD

- 영블록을 인수로 하는 곱을 없애보자.

$$A = U\Sigma V^T = \left[\mathbb{u}_1 \quad \cdots \quad \mathbb{u}_k \quad \Big| \quad \mathbb{u}_{k+1} \quad \cdots \quad \mathbb{u}_m\right] \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O_{k\times(n-k)} \\ 0 & \cdots & \sigma_k & \\ \hline & O_{(m-k)\times k} & & O_{(m-k)\times(n-k)} \end{array}\right] \begin{bmatrix} \mathbb{v}_1^T \\ \vdots \\ \mathbb{v}_k^T \\ - \\ \mathbb{v}_{k+1}^T \\ \vdots \\ \mathbb{v}_n^T \end{bmatrix}$$

$$\Downarrow$$

$$A = U_1\Sigma_1 V_1^T = \left[\mathbb{u}_1 \quad \cdots \quad \mathbb{u}_k\right] \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbb{v}_1^T \\ \vdots \\ \mathbb{v}_k^T \end{bmatrix}$$

- $U_1 : m\times k$, $\Sigma_1 : k\times k$, $V_1^T : k\times n$

축소된 SVD

- A 의 축소된 특이값 확장(reduced singular expansion)
  - $A = U_1 \Sigma_1 V_1^T = [\mathbb{u}_1 \quad \cdots \quad \mathbb{u}_k] \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbb{v}_1^T \\ \vdots \\ \mathbb{v}_k^T \end{bmatrix} = \sigma_1 \mathbb{u}_1 \mathbb{v}_1^T + \cdots + \sigma_k \mathbb{u}_\mathrm{k} \mathbb{v}_k^T$
  - 이 결과는 모든 행렬에 적용된다.

축소된 SVD 예제

- 행렬 $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$의 축소된 특이값분해를 구하라.

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

$$= U_1 \Sigma V_1^T$$

$$= \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

LSA (Latent Semantic Analysis)

- LSA(LSI)는 document data의 숨겨진 의미(hidden concept)를 찾아내는 기법이다.
- LSA는 각각의 문서(document)와 단어(word)를 벡터로 표현한다. 벡터내에서 각각의 element는 **숨겨진 의미**가 될 것이다.

LSA (Latent Semantic Analysis)

d_1 : *Romeo and Juliet.*

d_2 : *Juliet: O happy dagger!*

d_3 : *Romeo died by dagger.*

d_4 : *"Live free or die", that's the New-Hampshire's motto.*

d_5 : *Did you know, New-Hampshire is in New-England.*

query :
- dies
- dagger

- 3번 문서는 쿼리에 대해서 1등이 될 것이다.
- 2번, 4번 문서는 그 다음이 될 것이다.
- 1번, 5번 문서는?
  - ✓ 사람들이 인식하기로는 문서 1번이 문서 5번 보다 주어진 쿼리에 더 맞는 문서이다.

컴퓨터도 이러한 추론 과정을 할 수 있을까? 즉 숨겨진 의미를 찾을 수 있을까?

LSA (Latent Semantic Analysis)

matrix A :

	d_1	d_2	d_3	d_4	d_5
romeo	1	0	1	0	0
juliet	1	1	0	0	0
happy	0	1	0	0	0
dagger	0	1	1	0	0
live	0	0	0	1	0
die	0	0	1	1	0
free	0	0	0	1	0
new-hampshire	0	0	0	1	1

matrix $\boldsymbol{B}$ $= A^T A$ doc-doc matrix

즉, 문서 i와 문서 j가 b개의 단어를 가지고 있으면 $B[i,j] = b$

matrix C $= AA^T$ term-term matrix

즉, 단어 i와 단어 j가 c 문서에서 함께 발생했으면 $C[i,j] = c$

LSA (Latent Semantic Analysis)

SVD 사용!

$A = S\Sigma U^T$, S는 B의 eigenvector이고, U는 C의 eigenvector이다.

$$\Sigma = \begin{bmatrix} 2.285 & 0 & 0 & 0 & 0 \\ 0 & 2.010 & 0 & 0 & 0 \\ 0 & 0 & 1.361 & 0 & 0 \\ 0 & 0 & 0 & 1.118 & 0 \\ 0 & 0 & 0 & 0 & 0.797 \end{bmatrix}$$

singular value

LSA (Latent Semantic Analysis)

Reduced SVD 사용!

$A_k = S_k \Sigma_{\mathrm{k}} U_k^T$, 모든 singular value를 사용할 수 없고, 작은 것들은 제외한다.

k개의 특이값만 남기는 것이다. 즉 k개의 "hidden concepts"만 남긴다.

LSA (Latent Semantic Analysis)

$$\Sigma_2 = \begin{bmatrix} 2.285 & 0 \\ 0 & 2.010 \end{bmatrix}$$

$$U_2^T = \begin{bmatrix} -0.311 & -0.407 & -0.594 & -0.603 & -0.143 \\ 0.363 & 0.541 & 0.200 & -0.695 & -0.229 \end{bmatrix}$$

$$\begin{array}{r} \textit{romeo} \\ \textit{juliet} \\ \textit{happy} \\ \textit{dagger} \\ \textit{live} \\ \textit{die} \\ \textit{free} \\ \textit{new-hampshire} \end{array} \quad S_2 = \begin{bmatrix} -0.396 & 0.280 \\ -0.314 & 0.450 \\ -0.178 & 0.269 \\ -0.438 & 0.369 \\ -0.264 & -0.346 \\ -0.524 & -0.246 \\ -0.264 & -0.346 \\ -0.326 & -0.460 \end{bmatrix}$$

LSA (Latent Semantic Analysis)

$S_2\Sigma_2$

$$romeo = \begin{bmatrix} -0.905 \\ 0.563 \end{bmatrix},\ juliet = \begin{bmatrix} -0.717 \\ 0.905 \end{bmatrix},\ happy = \begin{bmatrix} -0.407 \\ 0.541 \end{bmatrix},\ dagger = \begin{bmatrix} -1.001 \\ 0.742 \end{bmatrix},$$

$$live = \begin{bmatrix} -0.603 \\ -0.695 \end{bmatrix},\ die = \begin{bmatrix} -1.197 \\ -0.494 \end{bmatrix},\ free = \begin{bmatrix} -0.603 \\ -0.695 \end{bmatrix},\ new\text{-}hampshire = \begin{bmatrix} -0.745 \\ -0.925 \end{bmatrix},$$

$\Sigma_2 U_2^T$

$$d_1 = \begin{bmatrix} -0.711 \\ 0.730 \end{bmatrix},\ d_2 = \begin{bmatrix} -0.930 \\ 1.087 \end{bmatrix},\ d_3 = \begin{bmatrix} -1.357 \\ 0.402 \end{bmatrix},\ d_4 = \begin{bmatrix} -1.378 \\ -1.397 \end{bmatrix},\ d_5 = \begin{bmatrix} -0.327 \\ -0.460 \end{bmatrix}.$$

$$q = \frac{\begin{bmatrix} -1.197 \\ -0.494 \end{bmatrix} + \begin{bmatrix} -1.001 \\ 0.742 \end{bmatrix}}{2} = \begin{bmatrix} -1.099 \\ 0.124 \end{bmatrix}$$

cosine similarity $\dfrac{d_i \cdot q}{|d_i||q|}$

LSA (Latent Semantic Analysis)