Chap12

특이값 분해(Singular Value Decomposition)

12.1 로우-랭크(Low-rank) 행렬에 의한 행렬의 근사

12.1.1 로우-랭크 행렬의 이점

Low-rank Assumption은 *'어떤 행렬 M 을 M 보다 더 작은 $rank$ 를 가지는 행렬의 곱으로 표현이 가능하다'* 라는 것을 의미한다. 예를 들어, $rank$ 가 1인 행렬을 생각해 보자. 모든 행들은 1차원 공간에 놓여 있으며, $\{v\}$ 는 이 공간에 대한 기저라고 하자. 행렬의 모든 행은 v 의 스칼라배이다. u 를 이러한 스칼라배들의 벡터라고 하면, 이러한 행렬을 uv^T 로 나타낼 수 있다. 이러한 표현으로 작은 저장공간으로 행렬을 나타낼 수 있으므로 로우-랭크 행렬의 이점이라 할 수 있다. 즉, 아래의 식과 같이 랭크 1 인 $m \times n$ 행렬에 대해 단지 $m+n$ 개의 숫자들로 구성할 수 있다.

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}^T = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

마찬가지로, 랭크가 2인 행렬은 다음과 같이 표현할 수 있다.

$$\begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ - & v_2^T & - \end{bmatrix}$$

그러나, 실제 데이터에서는 행렬 대부분은 로우-랭크가 아니다. 하지만, 종종 로우-랭크의 근사행렬이 거의 원래 행렬 만큼이나 잘 동작한다고 한다.

이번 장에서는 주어진 행렬에 대한 최적의 랭크-k 행렬을 찾는 방법에 대해 알아 보도록 한다. 이때의 랭크-k 행렬은 주어진 행렬에 가장 가까운 행렬이다.

12.1.2 행렬의 $Norm$

주어진 행렬에 가장 가까운 랭크-k 행렬을 찾는 문제를 정의하기 위해 행렬들에 대한 거리를 정의하는 것이 필요하다. 벡터들의 경우, 거리는 $norm$ 으로 주어지고, $norm$ 은 내적으로 정의된다. $\mathbb{R}$ 상의 벡터들에 대한 내적은 도트곱이고, 벡터의 $norm$ 은 그 원소들의 제곱의 합의 제곱근이다.

행렬의 $norm$은 행렬 A 를 벡터로 해석함으로써 계산할 수 있다. $m \times n$ 행렬은 mn-벡터로 표현할 수 있다.

아래와 같이 행렬 A 의 $norm$을 계산하는 것을 *프로베니우스(Frobenius)* $norm$ 이라고 한다.

$$\|A\|_F = \sqrt{\sum_i \sum_j A[i,j]^2}$$

Lemma : A의 프로베니우스 $norm$ 의 제곱은 A 의 행들의 제곱의 합과 동일하다.

- Proof : A 는 $m \times n$ 행렬이라 하고, A 를 행벡터로 나타내자.

$$A = \begin{bmatrix} - & a_1 & - \\ & \vdots & \\ - & a_m & - \end{bmatrix}$$

- 각각의 행-i 에 대해, 다음과 같이 나타낼 수 있다.

$$\|a_i\|^2 = a_i[1]^2 + a_i[2]^2 + \cdots + a_i[n]^2$$

- 이 식을 프로베니우스 $norm$ 의 식에 대입하면 다음과 같다.

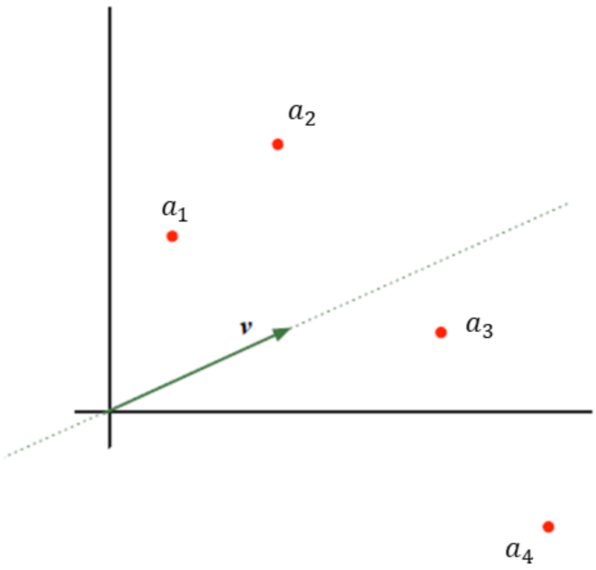
$$\begin{aligned} \|A\|_F &= (A[1,1]^2 + A[1,2]^2 + \cdots + A[1,n]^2) + \cdots + (A[m,1]^2 + A[m,2]^2 + \cdots + A[m,n]^2) \\ &= \|a_1\|^2 + \cdots + \|a_m\|^2 \end{aligned}$$

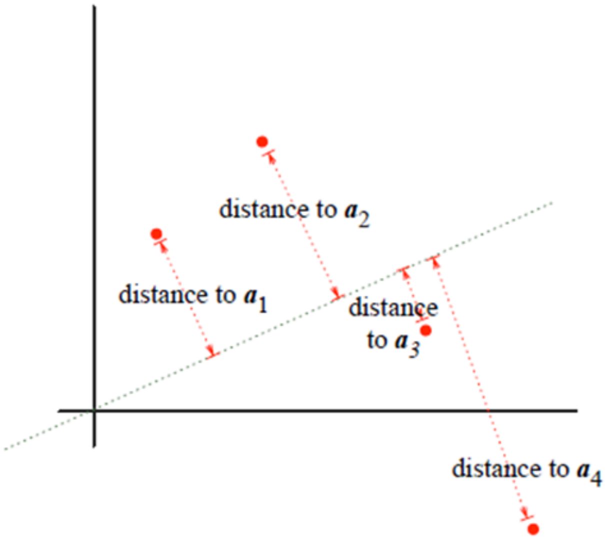
파이썬의 `numpy.linalg.norm()` 에서는 다양한 $norm$ 계산을 제공하는데 물론 프로베니우스 $norm$ 도 제공한다. 아래의 코드는 `numpy` 모듈을 이용한 프로베니우스 $norm$ 을 구하는 코드이다.

```
import numpy as np

A = np.matrix([[1, 0],
               [0, 1]])

fro_norm = np.linalg.norm(A, 'fro')
print(fro_norm)

'''출력결과
1.4142135623730951
'''
```

12.2 *트롤리 노선 위치 (Trolley-line-location)* 문제

9.1 소방차 문제 와 비슷하게 아래의 그림과 같이 벡터 $a_1, \ldots, a_m$ 으로 표현된 m개의 주택의 위치에 대해, 트롤리 노선을 어디에다 위치할 것인지를 찾는 문제를 생각해보자.





이 문제의 목적은 트롤리 노선을 m 개의 주택에 가능한 가깝게 배치하는 것이다.

각 벡터 a_i 는 트롤리 노선으로 부터 자신까지의 거리 d_i 를 가진다. 즉, 벡터 $[d_1, \ldots, d_m]$ 의 $norm$ 을 최소화 해햐한다. 이 문제는 벡터 $norm$ 의 제곱, $d_1^2 + \cdots + d_m^2$ 을 최소화하는 것과 동일하다.

12.2.1 트롤리 노선 위치 문제에 대한 솔루션

각 벡터 a_i 에 대해, $a_i = a_i^{||v} + a_i^{\perp v}$ 라고 나타낼 수 있으며 ($a_i^{||v}$ 는 v 에 따른 a_i 의 투영, $a_i^{\perp v}$ 는 v 에 직교하는 투영) 다음과 같이 쓸 수 있다. 단, 여기서 v는 $Span\{v\}$를 형성하는 단위 $norm$ 벡터($norm = 1$)이다.

$$a_1^{\perp v} = a_1 - a_1^{||v}$$

$$\vdots$$

$$a_m^{\perp v} = a_m - a_m^{||v}$$

피타고라스 정리에 의하면, 다음이 성립한다.

$$\left\|a_1^{\perp v}\right\|^2 = \|a_1\|^2 - \left\|a_1^{||v}\right\|^2$$

$$\vdots$$

$$\left\|a_m^{\perp v}\right\|^2 = \|a_m\|^2 - \left\|a_m^{||v}\right\|^2$$

$a_i^{\perp v}$ 는 a_i 에서 $Span\{v\}$ 까지의 거리이므로, 다음과 같이 나타낼 수 있다.

$$(a_1 에서 Span\{v\} 까지의 거리)^2 = \|a_1\|^2 - \left\|a_1^{||v}\right\|^2$$

$$\vdots$$

$$(a_m\text{에서 }Span\{v\}\text{까지의거리})^2 = \|a_m\|^2 - \left\|a_m^{||v}\right\|^2$$

위의 식을 수직으로 더하면 다음과 같다.

$$\begin{aligned}\sum_i (a_i\text{에서 }Span\{v\}\text{까지의거리})^2 &= \|a_1\|^2 + \cdots + \|a_m\|^2 - \left(\left\|a_1^{||v}\right\|^2 + \cdots + \left\|a_m^{||v}\right\|^2\right) \\ &= \|A\|_F^2 - \left(\left\|a_1^{||v}\right\|^2 + \cdots + \left\|a_m^{||v}\right\|^2\right)\end{aligned}$$

위의 12.1.2의 *Lemma* 에 의하면 A는 행들이 $a_1, \ldots, a_m$ 인 행렬이다.

v 는 $norm$이 1인 벡터이며, $a_i^{||v} = \langle a_i, v\rangle$ 이므로 $\left\|ai^{||v}\right\|^2 = \langle a_i, v\rangle^2$ 이다. 이를 이용하여 다음과 같이 나타낼 수 있다.

$$\sum_i (a_i\text{에서 }Span\{v\}\text{까지의거리})^2 = \|A\|_F^2 - \left(\langle a_1, v\rangle^2 + \langle a_2, v\rangle^2 + \cdots + \langle a_m, v\rangle^2\right)$$

$\left(\langle a_1, v\rangle^2 + \langle a_2, v\rangle^2 + \cdots + \langle a_m, v\rangle^2\right)$ 는 아래와 같이 $\|Av\|^2$ 로 나타낼 수 있다.

$$\begin{bmatrix} - & a_1 & - \\ & \vdots & \\ - & a_m & - \end{bmatrix} \begin{bmatrix} \\ v \\ \\ \end{bmatrix} = \begin{bmatrix} \langle a_1, v\rangle \\ \vdots \\ \langle a_m, v\rangle \end{bmatrix}$$

$$\|Av\|^2 = \left(\langle a_1, v\rangle^2 + \langle a_2, v\rangle^2 + \cdots + \langle a_m, v\rangle^2\right)$$

최종적으로 아래와 같은 식을 구할 수 있다.

$$\sum_i (a_i\text{에서 }Span\{v\}\text{까지의거리})^2 = \|A\|_F^2 - \|Av\|^2$$

따라서, 최적의 벡터 v 는 $\|Av\|^2$ 를 최대화하는 (즉, $\|Av\|$ 를 최대화하는) 단위벡터이다.

이를 이용하여 트롤리 노선문제에 대한 솔루션은 다음과 같이 pseudo-code로 나타낼 수 있다.

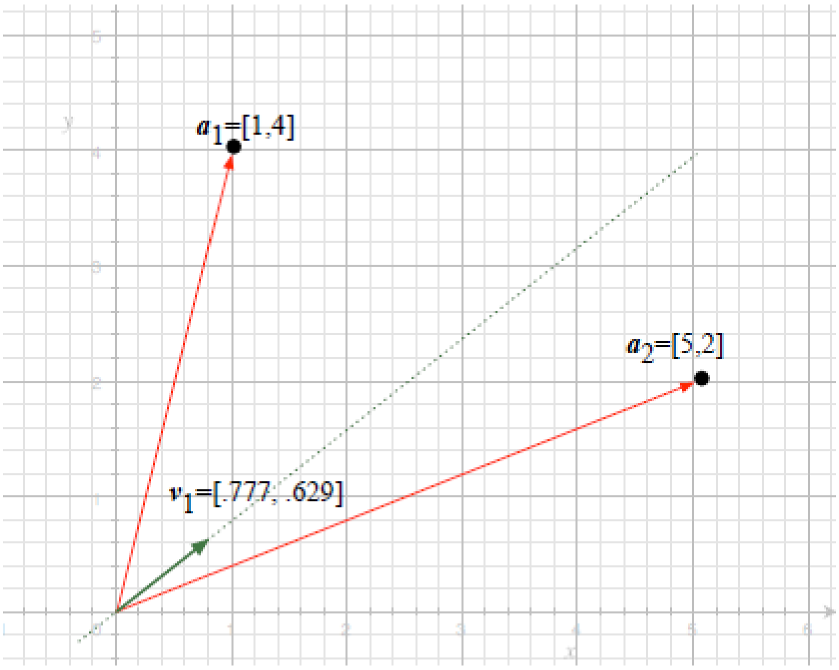
트롤리 노선 문제 솔루션

def trolley_line_location(A):
    Given a matrix A, find the vector v_1
    minimizing $\sum_i (a_i\text{에서 }Span\{v\}\text{까지의거리})^2$
    $v_1 = \arg\max\{\|Av\| : \|v\| = 1\}$
    $\sigma_1 = \|Av_1\|$
    return v_1

$\arg\max$ 표기법은 $\|Av\|$ 의 값이 최대가 되게 하는 것(이 경우 $norm$ 이 1인 벡터 v)을 의미한다. 이번 장에서는 솔루션을 제시하고 13장에서 v_1을 근사하는 방법에 대해 공부한다.

Definition : σ_1 은 A의 *첫 번째 특이값(singular value)* 이라 하고, v_1 은 *첫 번째 오른쪽 특이벡터* 라 한다.

Example 12.2.3 : $A = \begin{bmatrix} 1 & 4 \\ 5 & 2 \end{bmatrix}$ 라고 하면, 행렬 A의 행벡터 a_1, a_2는 $a_1 = [1, 4], a_2 = [5, 2]$ 이다. 이 경우, $\|Av\|$ 을 최대가 되게하는 단위벡터는 $v_1 \approx \begin{bmatrix} 0.78 \\ 0.63 \end{bmatrix}$ 이다. $\sigma_1 = \|Av\|$ 은 6.1 이다.



Theorem : A 는 $\mathbb{R}$ 상의 $m \times n$ 행렬이고, 이 행렬의 행들은 $a_1, \dots, a_m$ 이라 하자. 이때 v_1 은 A 의 첫 번째 오른쪽 특이벡터라고 하면, $Span\{v_1\}$ 은 다음을 최소화하는 1차원 벡터공간 $\mathcal{V}$ 이다.

$$(a_1 \text{에서 } \mathcal{V}\text{까지의 거리})^2 + \cdots + (a_m \text{에서 } \mathcal{V}\text{까지의 거리})^2$$

Lemma : 제곱 거리(squared distances)의 합의 최소값은 $\|A\|_F^2 - \|Av\|^2$ 이다.

12.2.3 생략

12.2.3 최상의 랭크-1 근사

위의 트롤리 노선 위치 문제에 대한 솔루션을 기반으로 하여 *주어진 행렬에 대한 최상의 랭크-1 근사를 찾는 문제* 에 대한 솔루션을 구해보자. 먼저, *랭크-1 근사* 문제를 정의하자.

Rank-one approximation

- *input* : 영이 아닌 행렬 A
- *output* : 프로베니우스 $norm$ 에 따라 A 에 가장 가까운 랭크-1 행렬 $\tilde{A}$

목적은 $\|A - \tilde{A}\|_F$ 을 최소화하는 랭크-1 행렬 $\tilde{A}$를 찾는 것이다.

$$\tilde{A} = \arg\min\{\|A - B\|_F : B\text{는 랭크 } 1\text{을 가진다}\}$$

어떤 랭크-1 행렬 $\tilde{A}$ 가 있다고 하면, 이 행렬은 A 에 얼마나 가까운가? A 와 $\tilde{A}$ 사이의 제곱 거리를 다음과 같이 나타낼 수 있다.

$$\left\|A-\tilde{A}\right\|_F^2=\left\|A-\tilde{A}\text{의 행 }1\right\|^2+\cdots+\left\|A-\tilde{A}\text{의 행 }m\right\|^2$$

위의 식이 의미하는 바는, A에 대한 거리를 최소화하기 위해서는 $\tilde{A}$ 의 각 행이 A 의 대응하는 행에 가능한한 가깝도록 되게 선택되어야 한다는 것을 말한다. 또한, $\tilde{A}$는 랭크가 1 이어야 한다. 즉, 어떤 벡터 v 에 대해, $\tilde{A}$의 각 행은 $Span\{v\}$ 내에 있어야 한다.

$$\tilde{A}=\begin{bmatrix} - & a_1\text{에 가장 가까운 }Span\{v\}\text{내의 벡터} & - \\ & \vdots & \\ - & a_m\text{에 가장 가까운 }Span\{v\}\text{내의 벡터} & - \end{bmatrix}$$

따라서, $i=1,\ldots,m$ 에대해, 다음과 같이 나타낼 수 있다.

$$\left\|A-\tilde{A}\text{의 행 }i\right\|=a_i\text{에서 }Span\{v\}\text{까지의 거리}$$

$$\left\|A-\tilde{A}\right\|^2=(a_1\text{에서 }Span\{v\}\text{까지의 거리})^2+\cdots+(a_m\text{에서 }Span\{v\}\text{까지의 거리})^2$$

12.2.4 최상의 랭크-1 근사에 대한 표현

위의 식을 표현하는 더 나은 방법이 있다. v를 *첫번째 오른쪽 특이벡터* v_1 이라 하고 a_i 에 가장 가까운 $Span\{v_1\}$의 벡터는 $a_i^{||v_1}$ 이며, 이것은 a_i의 $Span\{v_1\}$ 상으로의 투영이다. 식 $a_i^{||v_1}=\langle a_i,v_1\rangle\, v_1$ 을 사용하여 다음과 같이 나타낼 수 있다.

$$\tilde{A}=\begin{bmatrix} - & \langle a_1,v_1\rangle\, v_1^T & - \\ & \vdots & \\ - & \langle a_m,v_1\rangle\, v_1^T & - \end{bmatrix}$$

벡터-행렬 곱셈의 선형결합 해석을 사용하여 두 벡터의 외적으로 나타낼 수 있다.

$$\tilde{A}=\begin{bmatrix} \langle a_1,v\rangle \\ \vdots \\ \langle a_m,v\rangle \end{bmatrix}\begin{bmatrix} & v_1^T & \end{bmatrix}$$

$$\tilde{A}=\begin{bmatrix} \\ Av_1 \\ \\ \end{bmatrix}\begin{bmatrix} & v_1^T & \end{bmatrix}$$

$\sigma_1=\|Av_1\|$ (첫번째 왼쪽 특이값) 라고 하고, u_1은 $\sigma_1 u_1=Av_1$ 을 만족하는 $norm$-1 벡터라고 하면 다음과 같이 쓸 수 있다.

$$\tilde{A} = \sigma_1 \begin{bmatrix} \\ u_1 \\ \\ \end{bmatrix} \begin{bmatrix} & v_1^T & \end{bmatrix}$$

Definition : A 의 *첫 번째 왼쪽 특이벡터(left singular vector)* 는 $\sigma_1 u_1 = A v_1$ 을 만족하는 벡터 u_1 이다. 여기서, σ_1 은 A의 첫 번째 특이값이고 v_1은 첫 번째 오른쪽 특이벡터이다.

Theorem : A에 대한 최상의 랭크-1 근사는 $\sigma_1 u_1 v_1^T$ 이다. 여기서 여기서, σ_1 은 A의 첫 번째 특이값이고, u_1 은 첫 번째 왼쪽 특이벡터이며, v_1은 첫 번째 오른쪽 특이벡터이다.

Example 12.2.12 : Example 12.2.3 에서 보았듯이, 행렬 $A = \begin{bmatrix} 1 & 4 \\ 5 & 2 \end{bmatrix}$에 대해, 첫 번째 오른쪽 특이벡터는 $v_1 \approx \begin{bmatrix} 0.78 \\ 0.63 \end{bmatrix}$ 이고, 첫 번째 특이값 σ_1 은 약 6.1이다. 첫 번째 왼쪽 특이벡터는 $u_1 \approx \begin{bmatrix} 0.54 \\ 0.84 \end{bmatrix}$ 이며, 이것은 $\sigma_1 u_1 = A v_1$ 임을 의미한다.

$$\begin{aligned} \tilde{A} &= \sigma_1 u_1 v_1^T \\ &\approx 6.1 \begin{bmatrix} 0.54 \\ 0.84 \end{bmatrix} \begin{bmatrix} 0.78 & 0.63 \end{bmatrix} \\ &\approx \begin{bmatrix} 2.6 & 2.1 \\ 4.0 & 3.2 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} A - \tilde{A} &= \begin{bmatrix} 1 & 4 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 2.6 & 2.1 \\ 4.0 & 3.2 \end{bmatrix} \\ &\approx \begin{bmatrix} -1.56 & 1.93 \\ 1.00 & -1.23 \end{bmatrix} \end{aligned}$$

따라서, $A - \tilde{A}$ 의 제곱 프로베니우스 $norm$은 아래와 같다.

$$1.56^2 + 1.93^2 + 1^2 + 1.23^2 \approx 8.7$$

12.2.5 가장 가까운 1차원 아핀공간

12.2에서 트롤리 노선 문제를 정의할 때, 트롤리 노선은 원점을 지나간다고 규정하였다. 이러한 규정을 정의한 이유는 트롤리 노선 문제를 가장 가까운 1차원 *벡터공간* 을 찾는 문제에 대응시키기 위함이었다. 이러한 1차원 벡터공간은 원점을 지나는 직선이다. 4.5에서 배운 *아핀공간* 정의에 의하면 임의의 직선(반드시 원점을 지날 필요는 없음)은 *아핀공간(Affine space)* 이다.

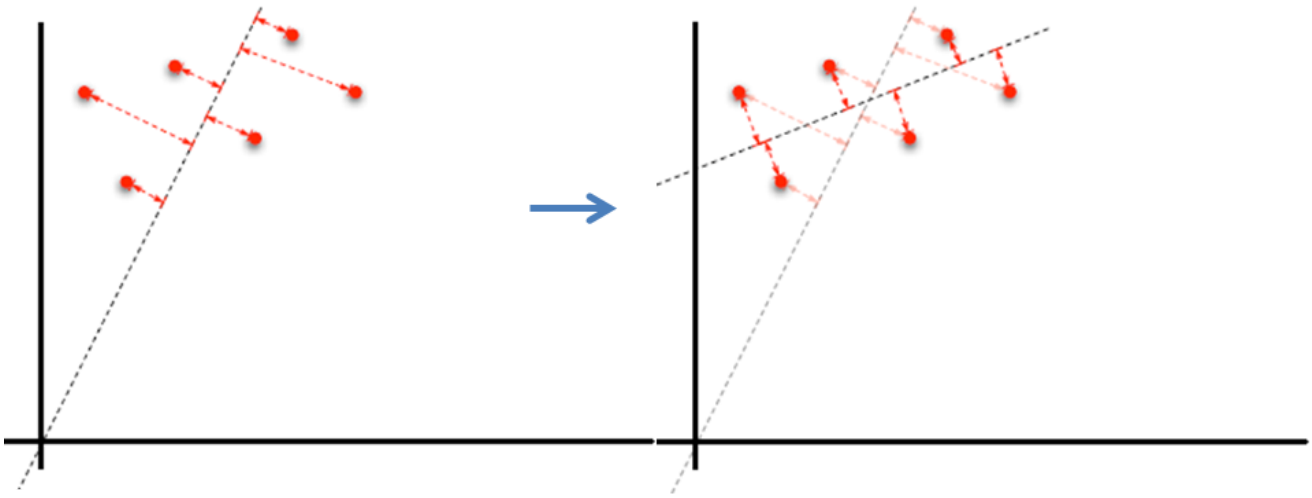
트롤리 노선 문제 솔루션을 적용하여 *가장 가까운 1차원 아핀공간* 을 찾을 수 있다. 주어진 점 $a_1, \ldots, a_m$ 에 대해, 점 $\bar{a}$ 를 선택하교, 다음에 $\bar{a}$를 빼주어 각 입력 점들을 평행이동하면 다음과 같다.

$$a_1 - \bar{a}, \ldots, a_m - \bar{a}$$

이러한 평행이동된 점들에 가장 가까운 1차원 벡터공간을 찾고, 그 다음에 $\bar{a}$ 를 더하여 찾은 벡터공간을 평행이동한다. 즉 주어진 점 $a_1, .., a_m$을 $\bar{a}$ 만큼 빼주어 평행이동 시킨다음 *원점을 지나는* 가장 가까운 1차원 벡터공간을 찾고, 그리고 다시 $\bar{a}$ 만큼 평행이동 한 것이 바로 가장 *가까운 아핀공간* 이 된다. 이러한 가장 가까운 아핀공간을 찾는것은 어떠한 $\bar{a}$ 를 설정하느냐에 따라 다르다. $\bar{a}$의 가장 최상의 선택은 입력 점들 $a_1, \ldots, a_m$의 *센트로이드(centroid)* 이다.

$$\bar{a} = \frac{1}{m}(a_1 + \cdots + a_m)$$

주어진 점들의 센트로이드를 찾은 다음, 그 센트로이드를 빼주어 주어진 점들을 평행이동하는 것을 점들에 대한 *센터링 (centering)* 이라고 한다.



12.3 가장 가까운 차원 - k 벡터공간

12.2의 트롤리 노선 위치 문제를 더 높은 차원으로 일반화하면 다음과 같다.

closest low-dimensional subspace :

- *input* : 벡터 $a_1, \ldots, a_m$ 과 양의 정수 k
- *output* : 다음을 최소화 하는 k-차원 벡터공간 $\mathcal{V}_k$에 대한 기저 $\rightarrow \sum_i (a_i \text{에서} \mathcal{V}_k \text{까지의 거리})^2$

트롤리 노선 문제는 $k = 1$ 인 단순히 특수한 경우이며, 1차원 벡터공간에 대한 기저를 찾는다.

12.3.1 특이값 및 특이벡터를 찾는 $Gedanken$ 알고리즘

이 알고리즘의 일반화는 *정규직교* 기저를 찾는 것이다. 이터레이션 i 에서, 선택된 벡터 v 는 $||Av||$ 을 최대가 되게 하는 것이며, 이 벡터는 이전에 선택된 모든 벡터들에 직교한다.

- v_1은 $||Av||$ 을 최대가 되게 하는 $norm$-1 벡터 v
- v_2는 v_1 에 직교하며 $||Av||$ 을 최대가 되게하는 $norm$-1 벡터 v
- v_3은 v_1 및 v_2에 직교하며 $||Av||$을 최대가 되게하는 $norm$-1 벡터 v
- $\vdots$

다음은 이 알고리즘에 대한 의사코드이다.

Given an $m \times n$ matrix A, find vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{\text{rank } A}$ such that, for $k = 1, 2, \ldots, \text{rank } (A)$, the k-dimensional subspace $\mathcal{V}$ that minimizes $\sum_i (\text{distance from row } i \text{ of } A \text{ to } \mathcal{V}_k)^2$ is Span $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$

def find_right_singular_vectors(A):
  for $i = 1, 2, \ldots$
    $\boldsymbol{v}_i = \arg\max\{\|A\boldsymbol{v}\| \ : \ \|\boldsymbol{v}\| = 1, \boldsymbol{v} \text{ is orthogonal to } \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \boldsymbol{v}_{i-1}\}$
    $\sigma_i = \|A\boldsymbol{v}_i\|$
  until $A\boldsymbol{v} = \mathbf{0}$ for every vector $\boldsymbol{v}$ orthogonal to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i$
  let r be the final value of the loop variable i.
  return $[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r]$

Definition : 벡터 $v_1, v_2, \ldots, v_r$ 은 A의 *오른쪽 특이벡터* 이고, 대응하는 실수 $\sigma_1, \sigma_2, \ldots, \sigma_r$ 은 A의 *특이값* 들이다.

12.3.2 특이값 및 오른쪽 특이벡터들의 성질

위의 12.3.1 에서 다음과 같은 성질을 알 수 있다.

Proposition : 오른쪽 특이벡터들은 *정규직교(orthonormal)* 한다.

Example 12.3.4 : Example 12.2.3 과 12.2.6 의 행렬 $A = \begin{bmatrix} 1 & 4 \\ 5 & 2 \end{bmatrix}$ 를 다시 보자. 첫 번째 오른쪽 특이벡터는 $v_1 \approx \begin{bmatrix} 0.78 \\ 0.63 \end{bmatrix}$ 이고, 첫 번째 특이값 $\sigma_1 \approx 6.1$ 이다. 두 번째 오른쪽 특이벡터는 $\begin{bmatrix} 0.78 \\ 0.63 \end{bmatrix}$ 에 직교하는 벡터들 중에서 선택되어야 한다. 따라서, 두번째 오른쪽 특이벡터는 $\begin{bmatrix} 0.63 \\ -0.78 \end{bmatrix}$ 이고 대응하는 특이값 $\sigma_2 \approx 2.9$ 이다.

벡터 v_1 과 v_2는 직교한다. 위의 예제에서 $\sigma_2 < \sigma_1$ 인 이유는 두 번째 최대화는 더 작은 후보 집합에 대해 수행되므로 첫 번째 값보다 더 클 수 없다.

벡터 v_1 과 v_2 는 직교하고 영이 아니므로, 두 벡터는 선형독립이고 $\mathbb{R}^2$ 를 생성한다.

Proposition : 특이값들은 음수가 아니며 *내림차순(descending order)* 이다.

- Proof : 각 특이값은 벡터의 $norm$이므로 음수가 될 수 없다. 각 $i > 1$ 에 대해, v_i 가 선택되는 벡터들의 집합은 v_{i-1} 이 선택된 벡터들의 집합의 부분집합이다. 따라서, 이터레이션 i에서의 최대값은 이터레이션 $i - 1$ 에서의 최대값 보다 클 수 없다. 즉, $\sigma_i \leq \sigma_{i-1}$ 이다.

Lemma : A 의 모든 행은 *오른쪽 특이벡터의 생성* 에 속한다.

- Proof : $\mathcal{V} = Span\{v_1, \ldots, v_r\}$ 이라 하자. $\mathcal{V}^o$는 $\mathcal{V}$ 의 소멸자(annihilator)라고 하고, $\mathcal{V}^o$는 $\mathcal{V}$에 직교하는 벡터들로 구성된다(10.6.4 참고). $\mathcal{V}^o$ 에 속하는 임의의 벡터 v^o 에 대해 곱 Av^o는 영벡터이다. 따라서, A의 행들은 v^o에 직교한다. 이 의미는 $(\mathcal{V}^o)^o = \mathcal{V}$ 이므로 A 의 행들이 $\mathcal{V}$에 속한다는 것을 보여준다.

12.3.3 특이값 분해

12.3.2의 Lemma에 따르면 A의 각 행 a_i 는 오른쪽 특이벡터들의 선형결합이다.

$$a_i = \sigma_{i1} v_1 + \cdots + \sigma_{ir} v_r$$

$v_1, \ldots, v_r$ 은 정규직교하므로, j 번째 항 $\sigma_{ij} v_j$는 j 번째 오른쪽 특이벡터 v_j에 따른 a_i의 투영이고 그 계수 σ_{ij} 는 a_i 와 v_j 의 내적이다.

$$a_i = \langle a_i, v_1 \rangle \, v_1 + \cdots + \langle a_i, v_r \rangle \, v_r$$

벡터-행렬 곱셈의 도트곱 정의를 사용하여 다음과 같이 표현할 수 있다.

$$a_i = \begin{bmatrix} \langle a_i, v_1 \rangle & \cdots & \langle a_i, v_r \rangle \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_r^T & - \end{bmatrix}$$

위의 식을 이용하여 행렬 A 를 행렬-행렬 곱으로 표현할 수 있다.

$$\begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} = \begin{bmatrix} \langle a_1, v_1 \rangle & \cdots & \langle a_1, v_r \rangle \\ \langle a_2, v_1 \rangle & \cdots & \langle a_2, v_r \rangle \\ & \vdots & \\ \langle a_m, v_1 \rangle & \cdots & \langle a_m, v_r \rangle \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_r^T & - \end{bmatrix}$$

위의 방정식을 더 간단하게 나타낼 수 있다. 우변의 첫 번째 행렬의 j 번째 열은 다음과 같다.

$$\begin{bmatrix} \langle a_1, v_j \rangle \\ \langle a_2, v_j \rangle \\ \vdots \\ \langle a_m, v_j \rangle \end{bmatrix}$$

이것은 선형결합의 도트곱 정의에 의해 Av_j 이다.

Definition : $\sigma_j u_j = Av_j$ 를 만족하는 벡터 $u_1, u_2, \ldots, u_r$ 은 A 의 *왼쪽 특이벡터(left singular vectors)* 이다.

Proposition : 왼쪽 특이벡터들은 정규직교이다.

왼쪽 특이벡터의 정의를 사용하여 $\sigma_j u_j$ 를 Av_j 에 대입하면 다음과 같다.

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} = \begin{bmatrix} | & & | \\ | & & | \\ \sigma_1 u_1 & \cdots & \sigma_r u_r \\ | & & | \\ | & & | \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_r^T & - \end{bmatrix}$$

마지막으로, $\sigma_1, \ldots, \sigma_r$ 을 대각행렬(diagonal matrix)로 분리하면 다음 방정식을 얻는다.

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} = \begin{bmatrix} | & & | \\ | & & | \\ u_1 & \cdots & u_r \\ | & & | \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_r^T & - \end{bmatrix}$$

Definition : 행렬 A 의 *특이값 분해* A 의 인수분해 $A = U\Sigma V^T$ 이다. 여기서, 행렬 U, Σ, V 는 다음 세 가지 성질을 가진다.

- Property S1 : Σ 는 대각행렬이고 그 원소들 $\sigma_1, \ldots, \sigma_r$ 은 양수이고 내림차순이다.
- Property S2 : V 는 열-직교행렬이다.
- Property S3 : U 는 열-직교행렬이다.

Theorem : $\mathbb{R}$ 상의 모든 행렬 A는 *특이값분해(SVD)* 를 가진다.

특이값 분해는 전치에 대해 대칭성이 있다. 행렬곱의 전치의 성질에 의하면 다음이 성립한다.

$$\begin{aligned} A^T &= (U\Sigma V^T)^T \\ &= V\Sigma^T U^T \\ &= V\Sigma U^T \end{aligned}$$

위의 식에서 Σ의 전치행렬은 Σ 자신이다. 따라서, A^T 의 SVD는 A의 SVD에서 U 와 V 를 바꾸면 된다.

12.3.4 가장 가까운 k-차원 공간을 찾는 데 오른쪽 특이벡터 사용하기

Lemma : $v_1, \ldots, v_k$ 는 벡터공간 $\mathcal{V}$ 에 대한 정규직교 벡터 기저라고 하면,

$$(a_1 \text{에서 } \mathcal{V}\text{까지의 거리})^2 + \cdots + (a_m \text{에서 } \mathcal{V}\text{까지의 거리})^2$$

은 $\|A\|_F^2 - \|Av_1\|^2 - \|Av_2\|^2 - \cdots - \|Av_k\|^2$ 이다.

- Proof : 행렬 A의 각 벡터 a_i 에 대해, $a_i = a_i^{||\mathcal{V}} + a_i^{\perp\mathcal{V}}$ 라고 표현하자. 피타고라스 정리에 의하면, $\left\|a_i^{\perp\mathcal{V}}\right\|^2 = \|a_1\|^2 - \left\|a_1^{||\mathcal{V}}\right\|^2$ 이다. 그러므로 제곱 거리의 합은 다음과 같다.

$$\left(\|a_1\|^2 - \left\|a_1^{||\mathcal{V}}\right\|^2\right) + \cdots + \left(\|a_m\|^2 - \left\|a_m^{||\mathcal{V}}\right\|^2\right)$$

- 위의 식은 아래와 동일하다.

$$\left(\|a_1\|^2 + \cdots + \|a_m\|^2\right) - \left(\left\|a_1^{||\mathcal{V}}\right\|^2 + \cdots + \left\|a_m^{||\mathcal{V}}\right\|^2\right)$$

- 위의 식에서 첫 번째 합 $\|a_1\|^2 + \cdots + \|a_m\|^2$ 은 $\|A\|_F^2$ 와 동일하다. 두 번째 합은 다음과 같다.

$$\left\|a_1^{||\mathcal{V}}\right\|^2 + \cdots + \left\|a_m^{||\mathcal{V}}\right\|^2$$
$$= \left(\left\|a_1^{||v_1}\right\|^2 + \cdots + \left\|a_m^{||v_k}\right\|^2\right) + \cdots + \left(\left\|a_m^{||v_1}\right\|^2 + \cdots + \left\|a_m^{||v_k}\right\|^2\right)$$
$$= \left(\langle a_1, v_1\rangle^2 + \cdots + \langle a_1, v_k\rangle^2\right) + \cdots + \left(\langle a_m, v_1\rangle^2 + \cdots + \langle a_m, v_k\rangle^2\right)$$

- 위의 내적을 다시 정리하면 최종적으로 다음과 같다.

$$\left(\langle a_1, v_1\rangle^2 + \cdots + \langle a_1, v_k\rangle^2\right) + \cdots + \left(\langle a_m, v_1\rangle^2 + \cdots + \langle a_m, v_k\rangle^2\right) = \|Av_1\|^2 + \|Av_2\|^2 + \cdots + \|Av_k\|^2$$

Theorem : A 는 $m \times n$ 행렬이라 하고 $a_1, \ldots, a_m$ 은 이 행렬의 행이라 하자. $v_1, \ldots, v_r$ 은 이 행렬의 오른쪽 특이벡터들이라 하고, $\sigma_1, \ldots, \sigma_r$ 은 특이벡터들에 대응하는 특이값이라하자. 임의의 양의 정수 $k \leq r$ 에 대해, $Span\{v_1, \ldots, v_k\}$ 은 k-차원 벡터공간 $\mathcal{V}$ 이며 이것은 다음을 최소화한다.

$$(a_1 \text{에서 } \mathcal{V}\text{까지의 거리})^2 + \cdots + (a_m \text{에서 } \mathcal{V}\text{까지의 거리})^2$$

즉, 제곱 거리의 합의 최소값은 $\|A\|_F^2 - \|Av_1\|^2 - \|Av_2\|^2 - \cdots - \|Av_k\|^2 = \|A\|_F^2 - \sigma_1^2 - \sigma_2^2 - \cdots - \sigma_k^2$ 이다.

- Proof : 공간 $\mathcal{V} = Span\{v_1, \ldots, v_k\}$ 에 대한 제곱 거리의 합은 아래와 같다.

$$\|A\|_F^2 - \sigma_1^2 - \sigma_2^2 - \cdots - \sigma_k^2 \quad \longrightarrow (1)$$

- 이값이 최소값임을 증명하기 위해서는 임의의 다른 k-차원 벡터공간 $\mathcal{W}$ 가 더 작지 않은 제곱의 합이 된다는 것을 보여줘야 한다. 임의의 k-차원 벡터공간 $\mathcal{W}$ 는 정규직교 기저를 가지며, $w_1, \ldots, w_k$ 는 이러한 기저라고 하자. 이 기저 벡터들을 위의 Lemma에 적용하면 $a_1, \ldots, a_m$ 에서 $\mathcal{W}$ 까지의 제곱 거리들의 합을 구할 수 있다.

$$\|A\|_F^2 - \|Aw_1\|^2 - \|Aw_2\|^2 - \cdots - \|Aw_k\|^2 \quad \longrightarrow (2)$$

- $\mathcal{V}$ 가 가장 가깝다는 것을 보여주기 위해, 식 (2)의 값이 식 (1)의 값보다 작지 않다는 것을 아래와 같이 보여줘야한다.

$$\|Aw_1\|^2 + \|Aw_2\|^2 + \cdots + \|Aw_k\|^2 \leq \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2$$

- 간단하게 하기 위해, 열들이 $w_1, \ldots, w_k$ 인 행렬 다음과 같이 W를 사용하여 나타내자.

$$\|AW\|_F^2 = \|Aw_1\|^2 + \|Aw_2\|^2 + \cdots + \|Aw_k\|^2$$

- 행렬 A 는 $A = U\Sigma V^T$ 로 인수분해 될 수 있다. 여기서, U 와 V 는 열-직교이고 Σ 는 대각행렬이다. $\|AW\|_F^2 = \left\|U\Sigma V^T W\right\|_F^2$ 이다. 이제, $i = 1, \ldots, k$ 에 대해, v_i 의 $\mathcal{W}$ 상으로의 투영을 생각해 보자.

$$v_i = v_i^{||\mathcal{W}} + v_i^{\perp\mathcal{W}}$$

- 위의 식에서 우변의 두 벡터들은 직교하므로 피타고리스 정리에 의해 다음이 성립한다.

$$\|v_i\|^2 = \left\|v_i^{||\mathcal{W}}\right\|^2 + \left\|v_i^{\perp\mathcal{W}}\right\|^2$$

- $\left\|v_i^{\|\mathcal{W}}\right\|^2 \leq \|v_i\|^2 = 1$이다. x_i 는 $v_i^{\|\mathcal{W}}$ 의 $w_1, \ldots, w_k$ 에 대한 좌표표현이라고 하자. 그렇게 되면, $v_i^{\|\mathcal{W}} = Wx_i$ 이다. W는 열-직교이므로, $\|Wx_i\| = \|x_i\|$ 이므로, $\|x_i\|^2 \leq 1$ 이다.
- $w_1, \ldots, w_k$ 는 정규직교이므로 다음과 같이 쓸 수 있다.

$$x_i = [\langle v_1, w_1 \rangle, \langle v_2, w_2 \rangle, \ldots, \langle v_i, w_k \rangle]$$

- 이 벡터는 행렬-행렬 곱셈의 벡터-행렬 정의 및 벡터-행렬 곱셈의 도트곱 정의에 의해 V^TW의 행 i이다.

$$\begin{aligned}\left\|\Sigma V^T W\right\|^2 &= \sigma_1^2 \left\|V^T W \text{의 행 } 1\right\|^2 + \cdots + \sigma_k^2 \left\|V^T W \text{의 행 } m\right\|^2 \\ &= \sigma_1^2 \|x_1\|^2 + \cdots + \sigma_k^2 \|x_k\|^2 \\ &\leq \sigma_1^2 + \cdots + \sigma_k^2\end{aligned}$$

12.3.5 A 에 대한 최상의 랭크-k 근사

12.2.4 에서 보았듯이, A에 대한 최상의 랭크-1 근사는 $\sigma_1 u_1 v_1^T$ 이다. 그렇다면 이것을 일반화 해보자.

Theorem : $k \leq rankA$ 에 대해, A의 최상의 랭크-k 근사는 다음과 같다.

$$\tilde{A} = \sigma_1 u_1 v_1^T + \cdots + \sigma_k u_k v_k^T$$

이때, $\left\|A - \tilde{A}\right\|_F^2 = \|A\|_F^2 - \sigma_1^2 - \sigma_2^2 - \cdots - \sigma_k^2$ 이다.

- Proof :

$$\left\|A - \tilde{A}\right\|_F^2 = \left\|A - \tilde{A} \text{의 행 } 1\right\|^2 + \cdots + \left\|A - \tilde{A} \text{의 행 } m\right\|^2$$

- $\tilde{A}$ 가 k 보다 작거나 같은 랭크를 가지기 위해서는 차원 k 의 어떤 벡터공간 $\mathcal{V}$ 에 $\tilde{A}$ 의 모든 행이 속해야한다. 벡터공간 $\mathcal{V}$ 에 대한 $\tilde{A}$ 의 최상의 선택은 다음과 같다.

$$\tilde{A} = \begin{bmatrix} - & a_1 \text{에 가장 가까운 } \mathcal{V} \text{에 속하는 벡터} & - \\ & \vdots & \\ - & a_m \text{에 가장 가까운 } \mathcal{V} \text{에 속하는 벡터} & - \end{bmatrix}$$

$$\left\|A - \tilde{A}\right\|^2 = (a_1 \text{에서 } \mathcal{V} \text{까지의 거리})^2 + \cdots + (a_m \text{에서 } \mathcal{V} \text{까지의 거리})^2$$

- 12.3.4의 Theorem에 의하면, $\mathcal{V}$ 까지의 제곱 거리들의 합을 최소화하기 위해서는 $\mathcal{V}$ 가 첫 k 개 오른쪽 특이벡터들의 생성이 되어야 하고, 그 제곱 거리들의 합은 $\|A\|_F^2 - \sigma_1^2 - \sigma_2^2 - \cdots - \sigma_k^2$ 이다.
- $i = 1, \ldots, m$ 에 대해, a_i 에 가장 가까운 $\mathcal{V}$ 에 속하는 벡터는 a_i 의 $\mathcal{V}$ 상으로의 투영이고, 다음과 같이 쓸 수 있다.

$$\begin{aligned}a_i \text{의 } \mathcal{V} \text{상으로의 투영} &= a_i \text{의 } v_1 \text{에 따른 투영} + \cdots + a_i \text{의 } v_m \text{에 따른 투영} \\ &= \langle a_i, v_1 \rangle v_1 + \cdots + \langle a_i, v_k \rangle v_k\end{aligned}$$

- 위의 식을 $a_1, \ldots, a_m$ 에 대하여 행렬의 덧셈 정의를 사용하면 다음과 같다.

$$\begin{aligned}\tilde{A} &= \begin{bmatrix} - & \langle a_1, v_1 \rangle v_1 & - \\ & \vdots & \\ - & \langle a_m, v_1 \rangle v_1 & - \end{bmatrix} + \cdots + \begin{bmatrix} - & \langle a_1, v_k \rangle v_k & - \\ & \vdots & \\ - & \langle a_m, v_k \rangle v_k & - \end{bmatrix} \\ &= \sigma_1 \begin{bmatrix} \\ u_1 \\ \\ \end{bmatrix} \begin{bmatrix} & v_1 & \end{bmatrix} + \cdots + \sigma_k \begin{bmatrix} \\ u_k \\ \\ \end{bmatrix} \begin{bmatrix} & v_k & \end{bmatrix}\end{aligned}$$

12.3.6 최상의 랭크-k 근사에 대한 행렬 형태

위의 식에서 보았듯이, A 에 대한 최상의 랭크-k 근사는 k 개 랭크-1 행렬들의 합이다. 이를 다음과 같이 쓸 수 있다.

$$\tilde{A} = \begin{bmatrix} | & & | \\ | & & | \\ u_1 & \cdots & u_r \\ | & & | \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_r^T & - \end{bmatrix}$$

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

12.3.7 영이 아닌 특이값들의 개수는 rank A 이다.

12.3.2의 Lemma 에 의해, 오른쪽 특이벡터들의 개수 r 은 적어도 A 의 랭크다.

$k = \text{rank } A$ 라고 하면, 이 k 값에 대해, A 에 대한 최상의 랭크-k 근사는 A 자신이다. 행렬 A 의 SVD를 다시보자.

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} = \underbrace{\begin{bmatrix} | & & | \\ | & & | \\ u_1 & \cdots & u_r \\ | & & | \\ | & & | \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_r^T & - \end{bmatrix}}_{V^T}$$

A의 각 행은 $U\Sigma$ 와 V^T 의 곱에 대응하는 행이다. 따라서, 선형결합에 의해 A 의 각 행은 V^T 의 행들의 선형결합이다. 그리고, V^T 의 행들은 서로 직교하고 영이 아닌 선형독립이다. 또한, 행들의 랭크는 $rankA$ 이고 이들의 생성의 차원은 $rankA$ 이다. 따라서, $RowA$ 와 $RowV^T$ 는 같다.

이와 마찬가지로 $ColA = ColU$ 이다. 그 이유는 A 의 각 열은 U 와 ΣV^T 의 열과의 곱이고, $\dim ColA = rankA = \dim ColU$ 이기 때문이다.

Proposition : A의 특이값 분해 $U\Sigma V^T$ 에서, $ColU = ColA$ 이고 $RowV^T = RowA$ 이다.

12.3.8 ~ 12.3.9 생략

12.3.10 U 는 열-직교임을 증명

12.3.3 의 특이값 분해의 성질 Property S3 에 의하면, 왼쪽 특이벡터들의 행렬 U 는 *열-직교* 이다. 그렇다면, 이 성질을 증명해보자.

왼쪽 특이벡터들은 $norm$ 이 1이 된다. 이들이 서로 직교한다는 것을 보여야 한다. $k = 0, 1, 2, \ldots, r$ 에 대해, $u_1, \ldots, u_k$ 는 서로 직교하고 나머지 왼쪽 특이벡터들 $u_{k+1}, \ldots, u_r$ 과는 직교한다는 것을 귀납법(개별적 사실에서 일반적 원리를 이끌어내는 방법)으로 증명한다.

$$U^{\perp} = \begin{bmatrix} - & u_1^* & - \\ - & u_2^* & - \\ & \vdots & \\ - & u_r^* & - \end{bmatrix}$$

$$U^{\perp}AV = \begin{bmatrix} - & u_1^* & - \\ - & u_2^* & - \\ & \vdots & \\ - & u_r^* & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ | & | & & | \\ \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_r u_r \\ | & | & & | \\ | & | & & | \end{bmatrix}$$

위에서 $u_1, \ldots, u_k$는 서로 열-직교이므로 귀납 가설에 의하면, 모든 $i < j < k$ 에 대해 $\langle u_i, u_j \rangle = 0$ 이다. 이것은 $u_1^* = u_1, u_2^* = u_2, \ldots, u_k^* = u_k$ 를 의미한다. 따라서, 다음과 같이 쓸 수 있다.

$$U^{\perp}AV = \begin{bmatrix} - & u_1 & - \\ - & u_2 & - \\ & \vdots & \\ - & u_k & - \\ - & u_{k+1}^* & - \\ & \vdots & \\ - & u_r^* & - \end{bmatrix} \begin{bmatrix} | & | & & | & & | \\ | & | & & | & & | \\ | & | & & | & & | \\ \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_k u_k & \cdots & \sigma_r u_r \\ | & | & & | & & | \\ | & | & & | & & | \\ | & | & & | & & | \end{bmatrix}$$
$$= \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & & 0 \\ & & \ddots & & & & \\ 0 & 0 & & \sigma_k & \beta_{k+1} & \cdots & \beta_r \\ 0 & 0 & & 0 & & & \\ & & \vdots & & & ? & \\ 0 & 0 & & 0 & & & \end{bmatrix}$$

위의 식에서 물음표(?)는 귀납단계에서 상관없는 행렬을 의미한다. 귀납 단계는 u_k 가 그 뒤에 나오는 모든 벡터득 $u_{k+1}, \ldots, u_r$ 에 직교한다는 것을 보여줘야 한다. 행렬에서 값 $\beta_{k+1}, \ldots, \beta_r$ 은 u_k 와 $\sigma_{k+1} u_{k+1}, \ldots, \sigma_r u_r$ 의 내적이다. 이 값이 영임을 보여주고자 한다.

w는 행렬 $U^{\perp}AV$ 의 k 번째 행이라 하면, $w = [0, 0, \ldots, \sigma_k, \beta_{k+1}, \ldots, \beta_r]$ 이고, 다음과 같이 쓸 수 있다.

$$\|w\|^2 = \sigma_k^2 + \beta_{k+1}^2 + \cdots + \beta_r^2$$

V 는 열-직교행렬이므로, 아래와 같이 V에 의한 곱은 $norm$을 보존한다. 따라서 다음이 성립한다.

$$\|Vw\|^2 = \sigma_k^2 + \beta_{k+1}^2 + \cdots + \beta_r^2$$

w의 처음 $k-1$ 개 원소들은 영이므로 Vw 는 V 의 열들, $k, k+1, \ldots, r$ 의 선형결합이다. V 의 열들은 서로 직교하므로, Vw 는 V의 처음 $k-1$ 열들 $v_1, \ldots, v_{k-1}$ 각각에 직교한다. $(U^{\perp}AV)w$ 의 원소 k 는 $\sigma_k^2 + \beta_{k+1}^2 + \cdots + \beta_r^2$ 이며, 다음과 같이 나타낼 수 있다.

$$\left\|U^{\perp}AVw\right\|^2 \geq \sigma_k^2 + \beta_{k+1}^2 + \cdots + \beta_r^2$$

$U^{\perp}$ 는 열-직교행렬이므로, $U^{\perp}$ 에 의한 곱은 $norm$을 보존한다.

$$\|AVw\|^2 \geq \sigma_k^2 + \beta_{k+1}^2 + \cdots + \beta_r^2$$

따라서, 다음 식을 얻을 수 있다.

$$\frac{\|A(Vw)\|}{\|Vw\|} \geq \frac{\sigma_k^2 + \beta_{k+1}^2 + \cdots + \beta_r^2}{\sqrt{\sigma_k^2 + \beta_{k+1}^2 + \cdots + \beta_r^2}} = \sqrt{\sigma_k^2 + \beta_{k+1}^2 + \cdots + \beta_r^2}$$

v 는 Vw 를 정규화하여 얻어지는 벡터 즉, $v = \frac{1}{\|Vw\|}Vw$ 라 하면, v 는 $v_1, \ldots, v_{k-1}$ 에 직교하는 $norm = 1$인 벡터이다. 만약, $\beta_{k+1}, \ldots, \beta_r$ 중 하나라도 영이 아니면, $\|Av\|$ 는 σ_k 보다 크게 된다. 그렇게 되면, k 번째 특이벡터를 $v_1, \ldots, v_{k-1}$ 에 직교하는 단위-$norm$ 벡터들 중에서 $\|Av\|$ 을 최대가 되게 하는 벡터를 선택하는 것과 모순이 된다. 따라서, $\beta_{k+1} = 0, \ldots, \beta_r = 0$ 임이 증명된다.

12.4 특이값 분해 사용하기

12.4.1 최소제곱에 SVD 사용하기

지난 10.8.5에서 행렬 A의 QR 분해는 $\|Ax-b\|$ 를 최소화하는 벡터 $\hat{x}$ 를 찾는 *최소제곱* 문제를 푸는 데 사용할 수 있다는 것을 알아보았다. 하지만 QR 분해는 A의 열들이 일차독립일 경우에만 적용될 수 있는 문제가 있었다. 이번에는 특이값 분해를 이용해 A 의 열들이 선형독립이 아니어도 최소제곱 문제를 푸는 방법에 대해 살펴본다.

$$V^T\hat{x} = \Sigma^{-1}U^Tb$$

$$\Sigma V^T\hat{x} = U^Tb$$

$$U\Sigma V^T\hat{x} = UU^Tb$$

$$A\hat{x} = UU^Tb$$

UU^Tb 는 b 의 $ColU$ 상으로의 투영 $b^{||ColU}$ 이다. $ColU = ColA$ 이므로 UU^Tb 는 b의 $ColA$ 상으로의 투영이다.