머신러닝/딥러닝을 위한 수학

선형대수

특이값 분해

16강

딥러닝호형

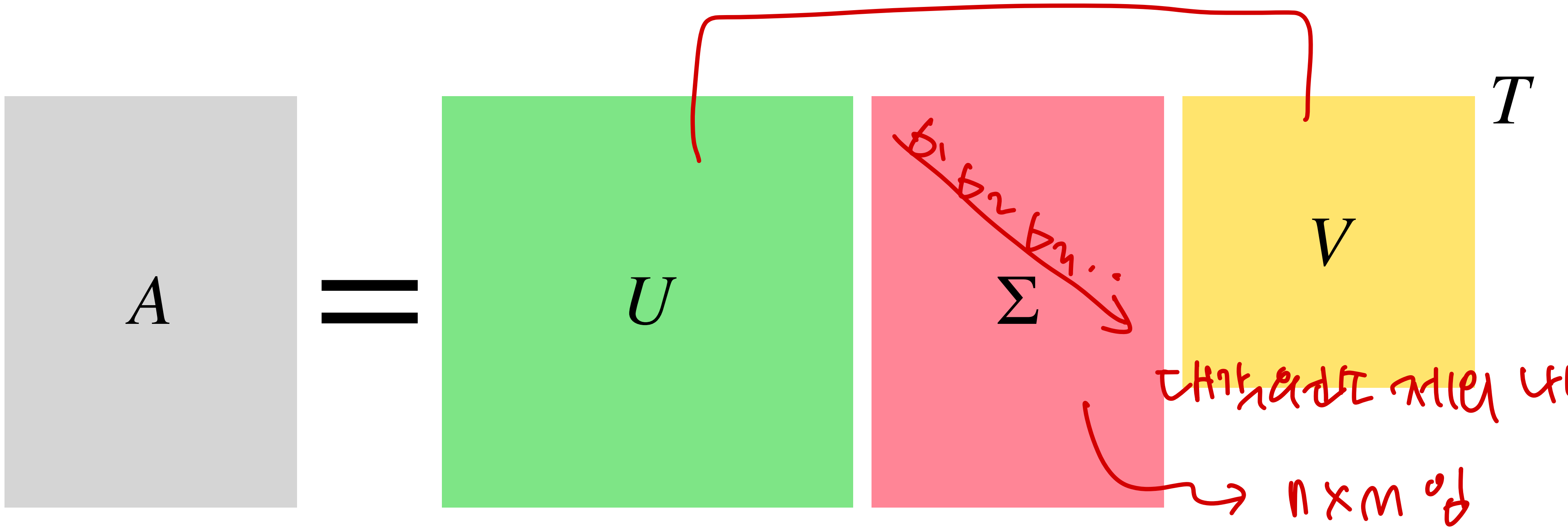
특이값 분해

앞서 대각화와 단원에서 $A = PDP^{-1}$ 임을 공부 → A가 $n \times n$인 정사각 행렬일 때만 가능.

특이값 분해 (SVD: Singular Value Decomposition)

A를 $m \times n$ 실수 행렬이라고 하면 orthogonal한 $m \times m$ 행렬 U, $n \times n$ 행렬 V가 존재해서 $A = U\Sigma V^T$로 나타낼 수 있다. 결론 이렇게 분해가능.

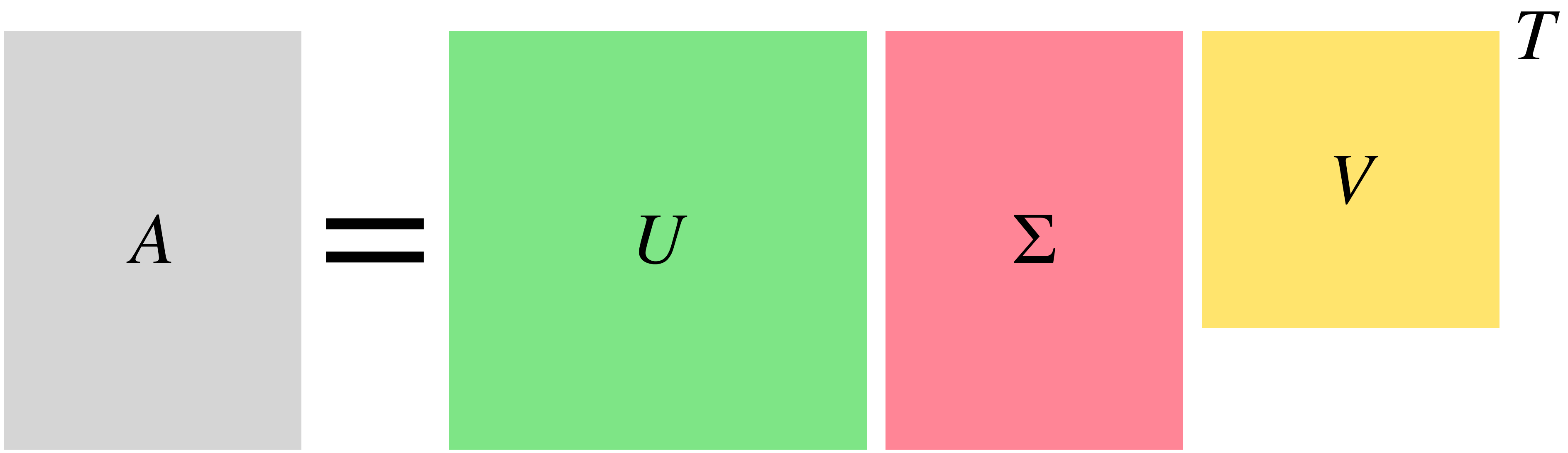
이 때, Σ는 $m \times n$ 행렬이고 대각 원소가 아닌 원소가 모두 0이고 대각 원소에 대해 $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_g$ $(g = min\{m, n\})$을 만족한다. 이미 알고있음



대각원소 제외 나머지 0 $\sigma_1 > \sigma_2 > \sigma_3 \cdots$

n×m 임

특이값 분해



1) $m < n$일 때

$$\Sigma = \left(\begin{array}{ccccc|c} \sigma_1 & 0 & 0 & 0 & 0 & \\ 0 & \sigma_2 & 0 & 0 & 0 & \\ 0 & 0 & \sigma_3 & 0 & 0 & O \\ 0 & 0 & 0 & \ddots & 0 & \\ 0 & 0 & 0 & 0 & \sigma_m & \end{array}\right)$$

2) $m > n$일 때

$$\Sigma = \left(\begin{array}{ccccc} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_n \\ \hline 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array}\right)$$

특이값 분해

rank $A = r$ 일 경우

$r < \min$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma_r & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

특이값 분해

U의 열벡터를 A의 좌 특이 벡터(left singular vectors), V의 열벡터를 A의 우 특이 벡터 (right singular vectors), Σ의 대각 성분을 A의 특이값(singular values)라고 한다.

정사각행렬

$AA^T = U\Lambda_u U^T$, $A^TA = V\Lambda_v V^T$이고 $\Sigma\Sigma^T = \Lambda_u$, $\Sigma^T\Sigma = \Lambda_v$라고 하면 정의하면

$\sigma_i = \sqrt{\lambda_i}$이다.($\because A^TAv = \lambda_i v \rightarrow v^TA^TAv = \lambda_i v^Tv \rightarrow \| Av \|^2 = \lambda_i \| v \|^2$)

고유벡터

즉, AA^T의 고유값은 항상 0보다 크거나 같다.

U에 대해 대각화하기 : U를 정사각행렬로 만들기. $A = m \times n$, $A^T = n \times m$

$\therefore AA^{-1} = m \times m$

$$A = U\Sigma V^T = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + \cdots + u_r\sigma_r v_r^T$$

σ_1 σ_2 σ_3

$= \sqrt{\lambda_i}$

얘가 대각선에 다 있는데 이 대각선 원소가 $\sqrt{\lambda_i}$임

A^TA와 AA^T의 고유값은 같음

특이값 분해

머신러닝/딥러닝: 어디서 사용되나요?

이미지 픽셀 = 행렬의 속성 → 이미지도 행렬분해가능

매우 다양한 곳에서 사용된다!!

특이값분해 $U\Sigma V^T$ 하면

$m \times n$ 이미지를 특이값 분해를 하면 다음과 같이 나타낼 수 있다.

가장 크게 영향 $\sigma_1 > \sigma_r$ ∴ 적은 시그마만 사용하면 시간↓

$$A = U\Sigma V^T = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + \cdots + u_r\sigma_r v_r^T$$

이 때 σ는 감소하게 배열되므로 $\sigma_1, \sigma_2, \cdots$ 순으로 값이 크다. 즉, 뒤로 갈수록 이미지에 관여하는 특성이 적으므로 적당한 $u_p\sigma_p v_p^T$까지만 사용한다.