Module 6.8 : Singular Value Decomposition

Let us get some more perspective on eigen vectors before moving ahead

- Let $v_1, v_2, \cdots, v_n$ be the eigen vectors of A and let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be corresponding eigen values

$$Av_1 = \lambda_1 v_1, Av_2 = \lambda_2 v_2, \cdots, Av_n = \lambda_n v_n$$

- Let $v_1, v_2, \cdots, v_n$ be the eigen vectors of A and let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be corresponding eigen values

$$Av_1 = \lambda_1 v_1, Av_2 = \lambda_2 v_2, \cdots, Av_n = \lambda_n v_n$$

- If a vector x in $\mathbb{R}^n$ is represented using $v_1, v_2, \cdots, v_n$ as basis then

$$x = \sum_{i=1}^{n} \alpha_i v_i$$

- Let $v_1, v_2, \cdots, v_n$ be the eigen vectors of A and let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be corresponding eigen values

$$Av_1 = \lambda_1 v_1, Av_2 = \lambda_2 v_2, \cdots, Av_n = \lambda_n v_n$$

- If a vector x in $\mathbb{R}^n$ is represented using $v_1, v_2, \cdots, v_n$ as basis then

$$x = \sum_{i=1}^{n} \alpha_i v_i$$

$$\text{Now, } Ax = \sum_{i=1}^{n} \alpha_i A v_i = \sum_{i=1}^{n} \alpha_i \lambda_i v_i$$

- Let $v_1, v_2, \cdots, v_n$ be the eigen vectors of A and let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be corresponding eigen values

$$Av_1 = \lambda_1 v_1, Av_2 = \lambda_2 v_2, \cdots, Av_n = \lambda_n v_n$$

- If a vector x in $\mathbb{R}^n$ is represented using $v_1, v_2, \cdots, v_n$ as basis then

$$x = \sum_{i=1}^{n} \alpha_i v_i$$

$$\text{Now, } Ax = \sum_{i=1}^{n} \alpha_i A v_i = \sum_{i=1}^{n} \alpha_i \lambda_i v_i$$

- The matrix multiplication reduces to a scalar multiplication if the eigen vectors of A are used as a basis.

- So far all the discussion was centered around square matrices $(A \in \mathbb{R}^{n \times n})$

- So far all the discussion was centered around square matrices ($A \in \mathbb{R}^{n \times n}$)
- What about rectangular matrices $A \in \mathbb{R}^{m \times n}$? Can they have eigen vectors?

- So far all the discussion was centered around square matrices ($A \in \mathbb{R}^{n\times n}$)
- What about rectangular matrices $A \in \mathbb{R}^{m\times n}$? Can they have eigen vectors?
- Is it possible to have $A_{m\times n} x_{n\times 1} = x_{n\times 1}$?

- So far all the discussion was centered around square matrices ($A \in \mathbb{R}^{n \times n}$)
- What about rectangular matrices $A \in \mathbb{R}^{m \times n}$? Can they have eigen vectors?
- Is it possible to have $A_{m \times n} x_{n \times 1} = x_{n \times 1}$? Not possible !

- So far all the discussion was centered around square matrices ($A \in \mathbb{R}^{n \times n}$)
- What about rectangular matrices $A \in \mathbb{R}^{m \times n}$? Can they have eigen vectors?
- Is it possible to have $A_{m \times n} x_{n \times 1} = x_{n \times 1}$? Not possible !
- The result of $A_{m \times n} x_{n \times 1}$ is a vector belonging to $\mathbb{R}^m$ (whereas $x \in \mathbb{R}^n$)

- So far all the discussion was centered around square matrices ($A \in \mathbb{R}^{n \times n}$)
- What about rectangular matrices $A \in \mathbb{R}^{m \times n}$? Can they have eigen vectors?
- Is it possible to have $A_{m \times n} x_{n \times 1} = x_{n \times 1}$? Not possible !
- The result of $A_{m \times n} x_{n \times 1}$ is a vector belonging to $\mathbb{R}^m$ (whereas $x \in \mathbb{R}^n$)
- So do we miss out on the advantage that a basis of eigen vectors provides for square matrices (i.e. converting matrix multiplications into scalar multiplications)?

- So far all the discussion was centered around square matrices ($A \in \mathbb{R}^{n \times n}$)
- What about rectangular matrices $A \in \mathbb{R}^{m \times n}$? Can they have eigen vectors?
- Is it possible to have $A_{m \times n} x_{n \times 1} = x_{n \times 1}$? Not possible !
- The result of $A_{m \times n} x_{n \times 1}$ is a vector belonging to $\mathbb{R}^m$ (whereas $x \in \mathbb{R}^n$)
- So do we miss out on the advantage that a basis of eigen vectors provides for square matrices (i.e. converting matrix multiplications into scalar multiplications)?
- We will see the answer to this question over the next few slides

- Note that matrix $A_{m \times n}$ provides a transformation $\mathbb{R}^n \to \mathbb{R}^m$

- Note that matrix $A_{m\times n}$ provides a transformation $\mathbb{R}^n \to \mathbb{R}^m$
- What if we could have pairs of vectors $(v_1, u_1), (v_2, u_2), \cdots, (v_k, u_k)$ such that $v_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$ and $Av_i = \sigma_i u_i$

- Note that matrix $A_{m \times n}$ provides a transformation $\mathbb{R}^n \to \mathbb{R}^m$
- What if we could have pairs of vectors $(v_1, u_1), (v_2, u_2), \cdots, (v_k, u_k)$ such that $v_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$ and $Av_i = \sigma_i u_i$
- Further let's assume that $v_1, \cdots, v_k, \cdots, v_n$ are orthogonal & thus form a basis V in $\mathbb{R}^n$

- Note that matrix $A_{m \times n}$ provides a transformation $\mathbb{R}^n \to \mathbb{R}^m$
- What if we could have pairs of vectors $(v_1, u_1), (v_2, u_2), \cdots, (v_k, u_k)$ such that $v_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$ and $Av_i = \sigma_i u_i$
- Further let's assume that $v_1, \cdots, v_k, \cdots, v_n$ are orthogonal & thus form a basis V in $\mathbb{R}^n$
- Similarly let's assume that $u_1, \cdots, u_k, \cdots, u_m$ are orthogonal & thus form a basis U in $\mathbb{R}^m$

- Note that matrix $A_{m \times n}$ provides a transformation $\mathbb{R}^n \to \mathbb{R}^m$
- What if we could have pairs of vectors $(v_1, u_1), (v_2, u_2), \cdots, (v_k, u_k)$ such that $v_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$ and $Av_i = \sigma_i u_i$
- Further let's assume that $v_1, \cdots, v_k, \cdots, v_n$ are orthogonal & thus form a basis V in $\mathbb{R}^n$
- Similarly let's assume that $u_1, \cdots, u_k, \cdots, u_m$ are orthogonal & thus form a basis U in $\mathbb{R}^m$
- Now what if every vector $x \in \mathbb{R}^n$ is represented using the basis V

- Note that matrix $A_{m \times n}$ provides a transformation $\mathbb{R}^n \to \mathbb{R}^m$
- What if we could have pairs of vectors $(v_1, u_1), (v_2, u_2), \cdots, (v_k, u_k)$ such that $v_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$ and $Av_i = \sigma_i u_i$
- Further let's assume that $v_1, \cdots, v_k, \cdots, v_n$ are orthogonal & thus form a basis V in $\mathbb{R}^n$
- Similarly let's assume that $u_1, \cdots, u_k, \cdots, u_m$ are orthogonal & thus form a basis U in $\mathbb{R}^m$
- Now what if every vector $x \in \mathbb{R}^n$ is represented using the basis V

$$x = \sum_{i=1}^{k} \alpha_i v_i$$

[note we are using k instead of n ; will clarify this in a minute]

- Note that matrix $A_{m \times n}$ provides a transformation $\mathbb{R}^n \to \mathbb{R}^m$
- What if we could have pairs of vectors $(v_1, u_1), (v_2, u_2), \cdots, (v_k, u_k)$ such that $v_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$ and $Av_i = \sigma_i u_i$
- Further let's assume that $v_1, \cdots, v_k, \cdots, v_n$ are orthogonal & thus form a basis V in $\mathbb{R}^n$
- Similarly let's assume that $u_1, \cdots, u_k, \cdots, u_m$ are orthogonal & thus form a basis U in $\mathbb{R}^m$
- Now what if every vector $x \in \mathbb{R}^n$ is represented using the basis V

$$x = \sum_{i=1}^{k} \alpha_i v_i$$

[note we are using k instead of n ; will clarify this in a minute]

$$Ax = \sum_{i=1}^{k} \alpha_i A v_i$$

- Note that matrix $A_{m \times n}$ provides a transformation $\mathbb{R}^n \to \mathbb{R}^m$
- What if we could have pairs of vectors $(v_1, u_1), (v_2, u_2), \cdots, (v_k, u_k)$ such that $v_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$ and $Av_i = \sigma_i u_i$
- Further let's assume that $v_1, \cdots, v_k, \cdots, v_n$ are orthogonal & thus form a basis V in $\mathbb{R}^n$
- Similarly let's assume that $u_1, \cdots, u_k, \cdots, u_m$ are orthogonal & thus form a basis U in $\mathbb{R}^m$
- Now what if every vector $x \in \mathbb{R}^n$ is represented using the basis V

$$x = \sum_{i=1}^{k} \alpha_i v_i$$

[note we are using k instead of n ; will clarify this in a minute]

$$Ax = \sum_{i=1}^{k} \alpha_i A v_i$$

$$= \sum_{i=1}^{k} \alpha_i \sigma_i u_i$$

- Note that matrix $A_{m\times n}$ provides a transformation $\mathbb{R}^n \to \mathbb{R}^m$
- What if we could have pairs of vectors $(v_1, u_1), (v_2, u_2), \cdots, (v_k, u_k)$ such that $v_i \in \mathbb{R}^n$, $u_i \in \mathbb{R}^m$ and $Av_i = \sigma_i u_i$
- Further let's assume that $v_1, \cdots, v_k, \cdots, v_n$ are orthogonal & thus form a basis V in $\mathbb{R}^n$
- Similarly let's assume that $u_1, \cdots, u_k, \cdots, u_m$ are orthogonal & thus form a basis U in $\mathbb{R}^m$
- Now what if every vector $x \in \mathbb{R}^n$ is represented using the basis V

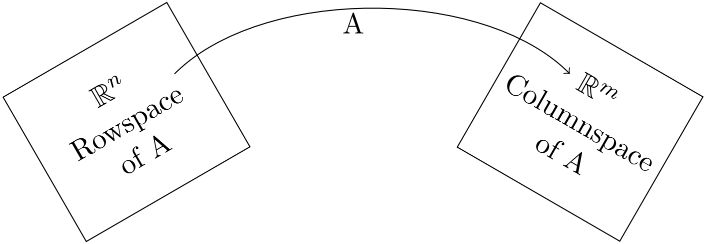
$$x = \sum_{i=1}^{k} \alpha_i v_i \qquad \text{[note we are using } k \text{ instead of } n \text{ ; will clarify this in a minute]}$$

$$Ax = \sum_{i=1}^{k} \alpha_i A v_i$$

$$= \sum_{i=1}^{k} \alpha_i \sigma_i u_i$$

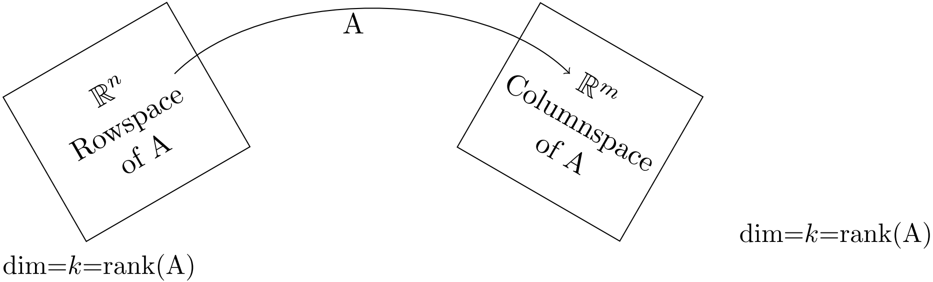
- Once again the matrix multiplication reduces to a scalar multiplication

Let's look at a geometric interpretation of this

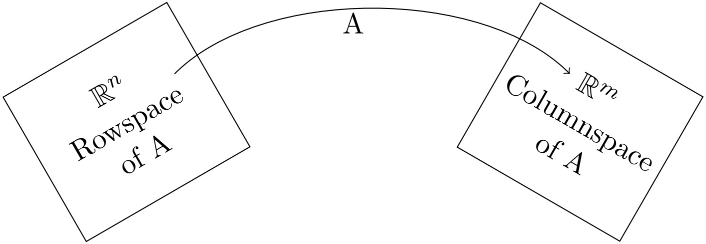


dim=k=rank(A)

dim=k=rank(A)



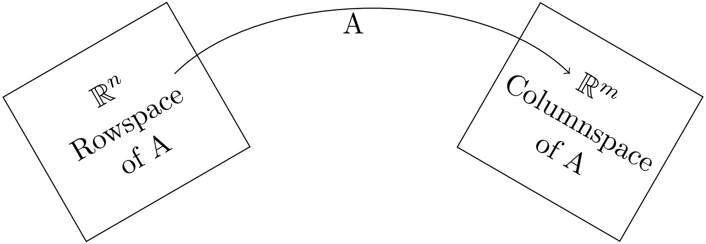
- $\mathbb{R}^n$ - Space of all vectors which can multiply with A to give Ax [this is the space of inputs of the function]



dim=k=rank(A)

dim=k=rank(A)

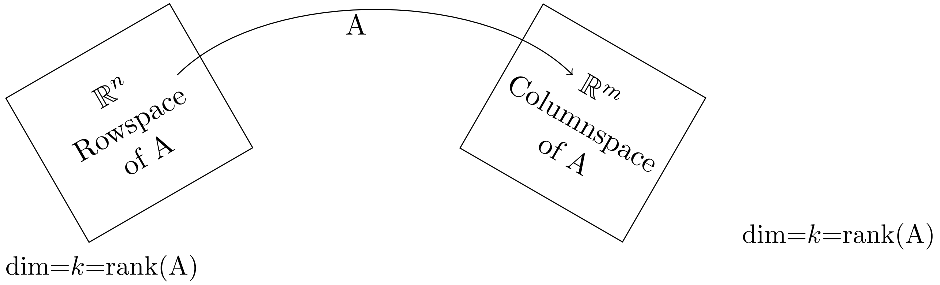
- $\mathbb{R}^n$ - Space of all vectors which can multiply with A to give Ax [this is the space of inputs of the function]
- $\mathbb{R}^m$ - Space of all vectors which are outputs of the function Ax



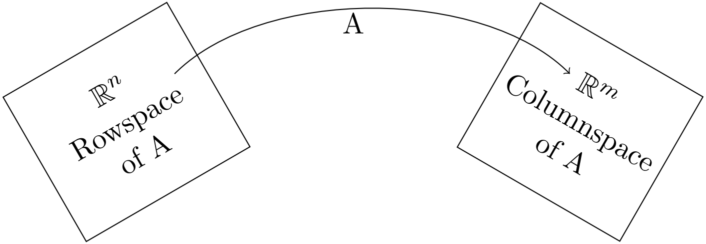
dim=k=rank(A)

dim=k=rank(A)

- $\mathbb{R}^n$ - Space of all vectors which can multiply with A to give Ax [this is the space of inputs of the function]
- $\mathbb{R}^m$ - Space of all vectors which are outputs of the function Ax
- We are interested in finding a basis U, V such that



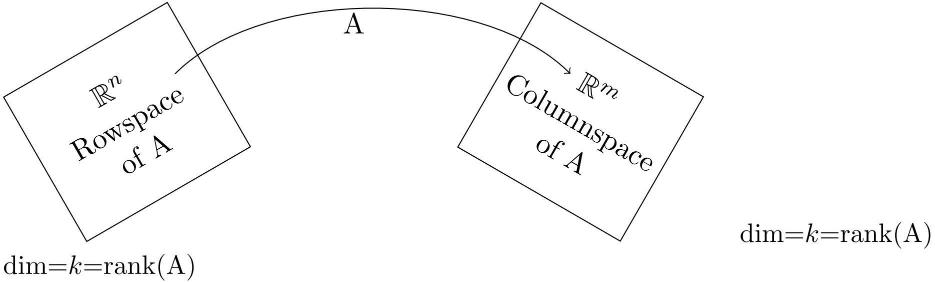
- $\mathbb{R}^n$ - Space of all vectors which can multiply with A to give Ax [this is the space of inputs of the function]
- $\mathbb{R}^m$ - Space of all vectors which are outputs of the function Ax
- We are interested in finding a basis U, V such that
  - V - basis for inputs



dim=k=rank(A)

dim=k=rank(A)

- $\mathbb{R}^n$ - Space of all vectors which can multiply with A to give Ax [this is the space of inputs of the function]
- $\mathbb{R}^m$ - Space of all vectors which are outputs of the function Ax
- We are interested in finding a basis U, V such that
  - V - basis for inputs
  - U - basis for outputs



- $\mathbb{R}^n$ - Space of all vectors which can multiply with A to give Ax [this is the space of inputs of the function]
- $\mathbb{R}^m$ - Space of all vectors which are outputs of the function Ax
- We are interested in finding a basis U, V such that
  - V - basis for inputs
  - U - basis for outputs
- such that if the inputs and outputs are represented using this basis then the operation Ax reduces to a scalar operation

- What do we mean by saying that dimension of rowspace is k? If $x \in \mathbb{R}^n$ then why is the dimension not n.

- What do we mean by saying that dimension of rowspace is k? If $x \in \mathbb{R}^n$ then why is the dimension not n.
- It means that of all the possible vectors in $\mathbb{R}^n$ only a subspace of vectors lying in $\mathbb{R}^k$ can act as inputs to Ax and produce a non-zero output. The remaining vectors in $\mathbb{R}^{n-k}$ will produce a zero output

- What do we mean by saying that dimension of rowspace is k? If $x \in \mathbb{R}^n$ then why is the dimension not n.
- It means that of all the possible vectors in $\mathbb{R}^n$ only a subspace of vectors lying in $\mathbb{R}^k$ can act as inputs to Ax and produce a non-zero output. The remaining vectors in $\mathbb{R}^{n-k}$ will produce a zero output
- Hence we need only k dimensions to represent x

$$x = \sum_{i=1}^{k} \alpha_i v_i$$

- Let's look at a way of writing this as a matrix operation

$$Av_1 = \sigma_1 u_1, Av_2 = \sigma_2 u_2, \cdots, Av_k = \sigma_k u_k$$

$$A_{m\times n} V_{n\times k} = U_{m\times k} \underbrace{\Sigma_{k\times k}}_{\text{diagonal matrix}}$$

- Let's look at a way of writing this as a matrix operation

$$Av_1 = \sigma_1 u_1, Av_2 = \sigma_2 u_2, \cdots, Av_k = \sigma_k u_k$$

$$A_{m\times n} V_{n\times k} = U_{m\times k} \underbrace{\Sigma_{k\times k}}_{\text{diagonal matrix}}$$

- If we have k orthogonal vectors ($V_{n\times k}$) then using Gram Schmidt orthogonalization, we can find $n-k$ more orthogonal vectors to complete the basis for $\mathbb{R}^n$ [We can do the same for U]

$$A_{m\times n} V_{n\times n} = U_{m\times m} \Sigma_{m\times n}$$

$$U^T A V = \Sigma \qquad [U^{-1} = U^T] \qquad A = U \Sigma V^T \qquad [V^{-1} = V^T]$$

- Let's look at a way of writing this as a matrix operation

$$Av_1 = \sigma_1 u_1, Av_2 = \sigma_2 u_2, \cdots, Av_k = \sigma_k u_k$$

$$A_{m\times n} V_{n\times k} = U_{m\times k} \underbrace{\Sigma_{k\times k}}_{\text{diagonal matrix}}$$

- If we have k orthogonal vectors ($V_{n\times k}$) then using Gram Schmidt orthogonalization, we can find $n-k$ more orthogonal vectors to complete the basis for $\mathbb{R}^n$ [We can do the same for U]

$$A_{m\times n} V_{n\times n} = U_{m\times m} \Sigma_{m\times n}$$

$$U^T A V = \Sigma \qquad [U^{-1} = U^T] \qquad A = U \Sigma V^T \qquad [V^{-1} = V^T]$$

- Σ is a diagonal matrix with only the first k diagonal elements as non-zero

- Let's look at a way of writing this as a matrix operation

$$Av_1 = \sigma_1 u_1, Av_2 = \sigma_2 u_2, \cdots, Av_k = \sigma_k u_k$$

$$A_{m\times n} V_{n\times k} = U_{m\times k} \underbrace{\Sigma_{k\times k}}_{\text{diagonal matrix}}$$

- If we have k orthogonal vectors ($V_{n\times k}$) then using Gram Schmidt orthogonalization, we can find $n-k$ more orthogonal vectors to complete the basis for $\mathbb{R}^n$ [We can do the same for U]

$$A_{m\times n} V_{n\times n} = U_{m\times m} \Sigma_{m\times n}$$

$$U^T AV = \Sigma \qquad [U^{-1} = U^T] \qquad A = U\Sigma V^T \qquad [V^{-1} = V^T]$$

- Σ is a diagonal matrix with only the first k diagonal elements as non-zero
- Now the question is how do we find V, U and Σ

- Suppose V, U and Σ exist, then

- Suppose V, U and Σ exist, then

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T)$$

- Suppose V, U and Σ exist, then

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

- Suppose V, U and Σ exist, then

$$
\begin{aligned}
A^T A &= (U\Sigma V^T)^T (U\Sigma V^T) \\
&= V\Sigma^T U^T U\Sigma V^T \\
A^T A &= V\Sigma^2 V^T
\end{aligned}
$$

- Suppose V, U and Σ exist, then

$$\begin{aligned} A^T A &= (U \Sigma V^T)^T (U \Sigma V^T) \\ &= V \Sigma^T U^T U \Sigma V^T \\ A^T A &= V \Sigma^2 V^T \end{aligned}$$

- What does this look like?

- Suppose V, U and Σ exist, then

$$\begin{aligned} A^T A &= (U\Sigma V^T)^T (U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ A^T A &= V\Sigma^2 V^T \end{aligned}$$

- What does this look like? Eigen Value decomposition of $A^T A$

- Suppose V, U and Σ exist, then

$$\begin{aligned} A^T A &= (U\Sigma V^T)^T (U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ A^T A &= V\Sigma^2 V^T \end{aligned}$$

- What does this look like? Eigen Value decomposition of $A^T A$
- Similarly we can show that

$$AA^T = U\Sigma^2 U^T$$

- Suppose V, U and Σ exist, then

$$\begin{aligned} A^TA &= (U\Sigma V^T)^T(U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ A^TA &= V\Sigma^2 V^T \end{aligned}$$

- What does this look like? Eigen Value decomposition of A^TA
- Similarly we can show that

$$AA^T = U\Sigma^2 U^T$$

- Thus U and V are the eigen vectors of AA^T and A^TA respectively and $\Sigma^2 = \Lambda$ where Λ is the diagonal matrix containing eigen values of A^TA

$$
\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}_{m \times n} = \begin{bmatrix} \uparrow & \cdots & \uparrow \\ u_1 & \cdots & u_k \\ \downarrow & \cdots & \downarrow \end{bmatrix}_{m \times k} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix}_{k \times k} \begin{bmatrix} \leftarrow & v_1 & \rightarrow \\ & \vdots & \\ \leftarrow & v_k & \rightarrow \end{bmatrix}_{k \times n}
$$

$$
= \sum_{i=1}^{k} \sigma_i u_i v_i^T
$$

$$
\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}_{m\times n} = \begin{bmatrix} \uparrow & \cdots & \uparrow \\ & & \\ u_1 & \cdots & u_k \\ \downarrow & \cdots & \downarrow \end{bmatrix}_{m\times k} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix}_{k\times k} \begin{bmatrix} \leftarrow & v_1 & \rightarrow \\ & \vdots & \\ \leftarrow & v_k & \rightarrow \end{bmatrix}_{k\times n}
$$

$$
= \sum_{i=1}^{k} \sigma_i u_i v_i^T
$$

Theorem:

$\sigma_1 u_1 v_1^T$ is the best rank-1 approximation of the matrix A. $\sum_{i=1}^{2} \sigma_i u_i v_i^T$ is the best rank-2 approximation of matrix A. In general, $\sum_{i=1}^{k} \sigma_i u_i v_i^T$ is the best rank-k approximation of matrix A. In other words, the solution to

$$
\min \|A - B\|_F^2 \quad \text{is given by :}
$$

$$
B = U_{.,k} \Sigma_{k,k} V_{k,.}^T \quad \text{(minimizes reconstruction error of A)}
$$

$$\sigma_i = \sqrt{\lambda_i} = \text{singular value of A}$$

$$\sigma_i = \sqrt{\lambda_i} = \text{singular value of A}$$
$$U = \text{left singular matrix of A}$$

$$\sigma_i = \sqrt{\lambda_i} = \text{singular value of A}$$
$$U = \text{left singular matrix of A}$$
$$V = \text{right singular matrix of A}$$