BITS Pilani
Pilani Campus

Mathematical Foundations for Data Science

MFDS Team

innovate achieve lead

BITS Pilani
Pilani Campus

DSECL ZC416, MFDS

Lecture No. 4

Agenda

- Linear Independence
- Inner product
- Gram Schmidt Orthogonalization Process
- Singular Value Decomposition
    - Principal Component Analysis
    - Dimensionality Reduction

Linear Independence

A set of vectors $\{v_1.........v_p\}$ in a vector space V is said to be linearly independent if the vector equation

$c_1v_1 + c_2v_2 + ...+c_pv_p = 0$ has $c_1 = 0,..........c_p = 0$ as the only solution

Eg) $\{(1,0), (0,1)\}$ is LI in R^2

$\{(1,0)\}$ is LI in R^2

•Inner Product Space

If u and v are vectors in R^n , then u and v are regarded as n x1 matrices. u^T is a 1 x n matrix. u^Tv is called Inner product of u and v is denoted by <u,v>

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in R^n and c is a scalar, then the following properties are true.

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
3. $c(\mathbf{u} \cdot \mathbf{v}) = (c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v})$
4. $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$
5. $\mathbf{v} \cdot \mathbf{v} \geq 0$, and $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

A vector space V with an inner product is called Inner Product Space. Whenever an inner product space is referred to, assume that set of scalars is the set of real numbers

Inner Product Space Example

An inner product on M_{2x2}

Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ be matrices in the vector space M_{2x2}

Define $\langle A,B \rangle = a_{11}b_{11} + a_{21}b_{21} + a_{12}\, b_{12} + a_{22}b_{22}$

It is easy to verify that the operator is an inner product

For eg: $\langle A,A \rangle = a_{11}^2 + a_{12}^2 + a^2_{21} + a_{22}^2$

≥ 0 always

$\langle A, A \rangle = 0 \Rightarrow a_{ij} = 0 \quad \forall\, i\, ,j$

$\Rightarrow A = 0$

Orthogonality

Orthogonal Set

Let V be vector space with an inner product

Non-zero vectors $v_1, v_2, \ldots\ldots\ldots, v_k \in V$ form an **orthogonal set** if they are orthogonal to each other:

$< v_i, v_j > = 0$ for $i \neq j$

If, in addition, all vectors are of unit norm $\|v_i\| = 1$ then $v_1, v_2, \ldots\ldots\ldots, v_k$ is called **orthonormal set**

Remark:

Any orthogonal set is linearly independent

Eg: {(1,1), (1,-1)} $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

Gram Schmidt Orthogonalization

Let V be a vector space with an inner product.
Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a basis for V. Let

$$\mathbf{v}_1 = \mathbf{x}_1,$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1,$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

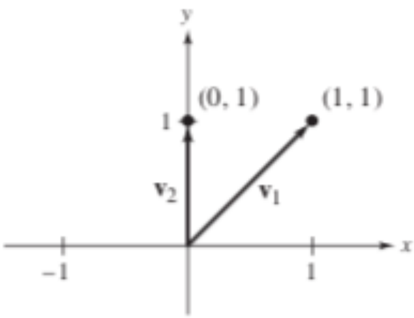
$$\mathbf{v}_n = \mathbf{x}_n - \frac{\langle \mathbf{x}_n, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \cdots - \frac{\langle \mathbf{x}_n, \mathbf{v}_{n-1} \rangle}{\langle \mathbf{v}_{n-1}, \mathbf{v}_{n-1} \rangle}\mathbf{v}_{n-1}.$$

Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthogonal basis for V.

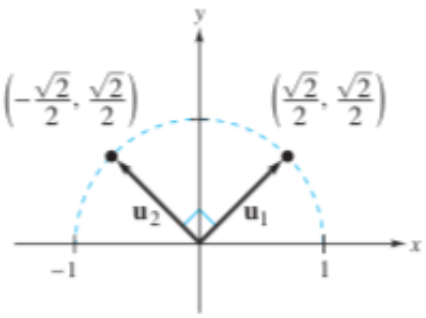
Gram Schmidt Orthonormalization

Apply the Gram-Schmidt orthonormalization process to the basis for R^2 shown below.

$$B = \{\overset{\mathbf{v}_1}{(1, 1)}, \overset{\mathbf{v}_2}{(0, 1)}\}$$



Given basis: $B = \{\mathbf{v}_1, \mathbf{v}_2\}$



Orthonormal basis: $B'' = \{\mathbf{u}_1, \mathbf{u}_2\}$

The Gram-Schmidt orthonormalization process produces

$$\mathbf{w}_1 = \mathbf{v}_1 = (1, 1)$$

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1$$

$$= (0, 1) - \tfrac{1}{2}(1, 1) = \left(-\tfrac{1}{2}, \tfrac{1}{2}\right).$$

The set $B' = \{\mathbf{w}_1, \mathbf{w}_2\}$ is an orthogonal basis for R^2. By normalizing each vector in B', you obtain

$$\mathbf{u}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} = \frac{1}{\sqrt{2}}(1, 1) = \frac{\sqrt{2}}{2}(1, 1) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$$

$$\mathbf{u}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \frac{1}{1/\sqrt{2}}\left(-\frac{1}{2}, \frac{1}{2}\right) = \sqrt{2}\left(-\frac{1}{2}, \frac{1}{2}\right) = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right).$$

So, $B'' = \{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for R^2.

QR Factorization

If A is an m x n matrix with linearly independent columns, then A can be factored as **A = QR** where Q is an m x n matrix whose columns form an orthonormal basis for Col A and **R is an n x n upper triangular invertible matrix** with positive entries on its diagonal

$$A_{mxn} = Q_{mxn} R_{nxn}$$

$$A = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ 0 & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{nn} \end{bmatrix}$$

vectors $q_1, \ldots, q_n$ are orthonormal m-vectors:

$$\|q_i\| = 1, \qquad q_i^T q_j = 0 \quad \text{if } i \neq j$$

QR Factorization Example

Find a QR factorization of A = $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

$A_{mxn} = Q_{mxn} . R_{nxn}$

1. First verify that the columns of A are LI

- Gram Schmidt Orthogonalization to the columns of A

$<v_1 .v_2 ..... v_n>$ such that $<v_i ,v_j> = 0$ for all $i \neq j$

- Normalize v_i to get u_i →columns of Q

- Use Q to get R

QR Factorization Example

$$
Q = \begin{array}{c} \\ \\ \end{array}
\begin{array}{ccc}
u_1 & u_2 & u_3 \\
\end{array}
$$

$$
\begin{array}{c} \\ Q = \end{array}
\begin{array}{c}
\begin{array}{ccc} \quad u_1 & \qquad u_2 & \qquad u_3 \end{array} \\
\begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix}
\end{array}
$$

To find R, observe that $Q^TQ = I$ because columns of Q are orthonormal

$$Q^TA = Q^T(QR) = IR = R$$

$$R = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{bmatrix}$$

LU Decomposition

LU Decomposition is an approach designed to exploit triangular systems for square matrices

We can write A = LU where L is a lower triangular matrix and U is an upper triangular matrix

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{bmatrix} = LU$$

$$\text{Where } L = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix} \text{ and } U = \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ 0 & U_{22} & U_{23} \\ 0 & 0 & U_{33} \end{bmatrix}$$

LU Decomposition

Multiplying out L and U and setting the answer equal to A

$$\begin{bmatrix} U_{11} & U_{12} & U_{13} \\ L_{21}U_{11} & L_{21}U_{12}+U_{22} & L_{21}U_{13}+U_{23} \\ L_{31}U_{11} & L_{31}U_{12}+L_{32}U_{22} & L_{31}U_{13}+L_{32}U_{23}+U_{33} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{bmatrix}$$

$$\boxed{U_{11}=1}, \quad \boxed{U_{12}=2}, \quad \boxed{U_{13}=4}$$

$$L_{21}U_{11}=3 \quad \therefore L_{21}\times 1=3 \quad \therefore \boxed{L_{21}=3},$$

$$L_{21}U_{12}+U_{22}=8 \quad \therefore 3\times 2+U_{22}=8 \quad \therefore \boxed{U_{22}=2},$$

$$L_{21}U_{13}+U_{23}=14 \quad \therefore 3\times 4+U_{23}=14 \quad \therefore \boxed{U_{23}=2}$$

LU Decomposition

$$L_{31}U_{11} = 2 \quad \therefore L_{31} \times 1 = 2 \quad \therefore \boxed{L_{31} = 2},$$

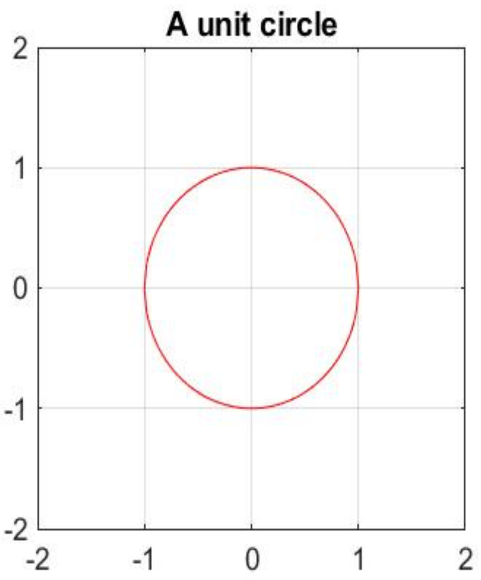
$$L_{31}U_{12} + L_{32}U_{22} = 6 \quad \therefore 2 \times 2 + L_{32} \times 2 = 6 \quad \therefore \boxed{L_{32} = 1},$$

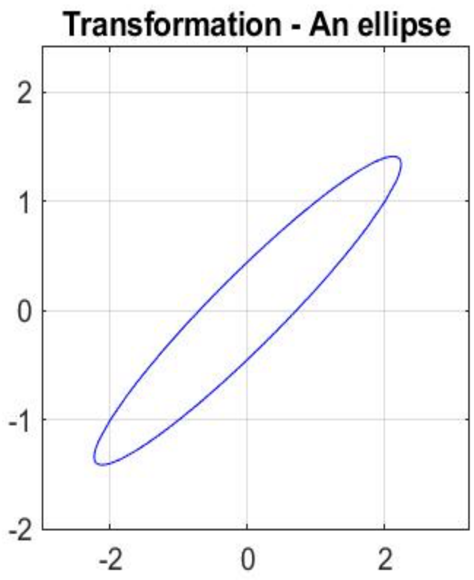
$$L_{31}U_{13} + L_{32}U_{23} + U_{33} = 13 \quad \therefore (2 \times 4) + (1 \times 2) + U_{33} = 13 \quad \therefore \boxed{U_{33} = 3}$$

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$$

This is LU Decomposition of A

Transformation of Circle under Matrix Operation





$Av_i = \sigma_i u_i$ where A is given matrix, v_i is orthogonal set
σ_i are singular values
u_i are principal axes direction

In higher dimensions
$Av_1 = \sigma_1 u_1$
$Av_2 = \sigma_2 u_2$
$Av_n = \sigma_n u_n$

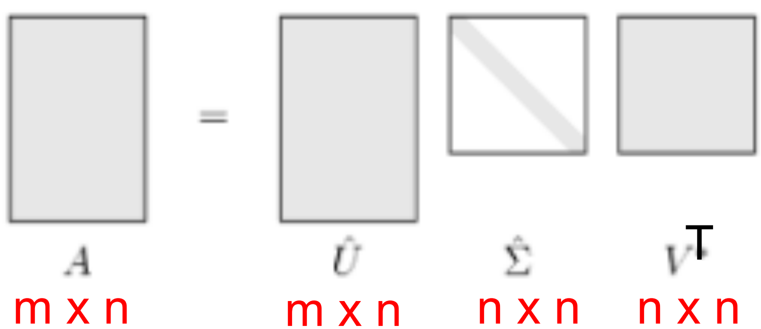
Reduced SVD

$Av_j = \sigma_j u_j \quad 1 \le j \le n$

$$\begin{bmatrix} \\ \\ & A & \\ \\ \\ \end{bmatrix} \begin{bmatrix} \\ v_1 \,\Big|\, v_2 \,\Big|\, \cdots \,\Big|\, v_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ u_1 \,\Big|\, u_2 \,\Big|\, \cdots \,\Big|\, u_n \\ \\ \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix},$$

$$A = \hat{U}\hat{\Sigma}V^T.$$

Reduced SVD ($m \ge n$)



Singular Value Decomposition

Singular Value Decomposition(SVD) is a factorization of an m x n matrix into

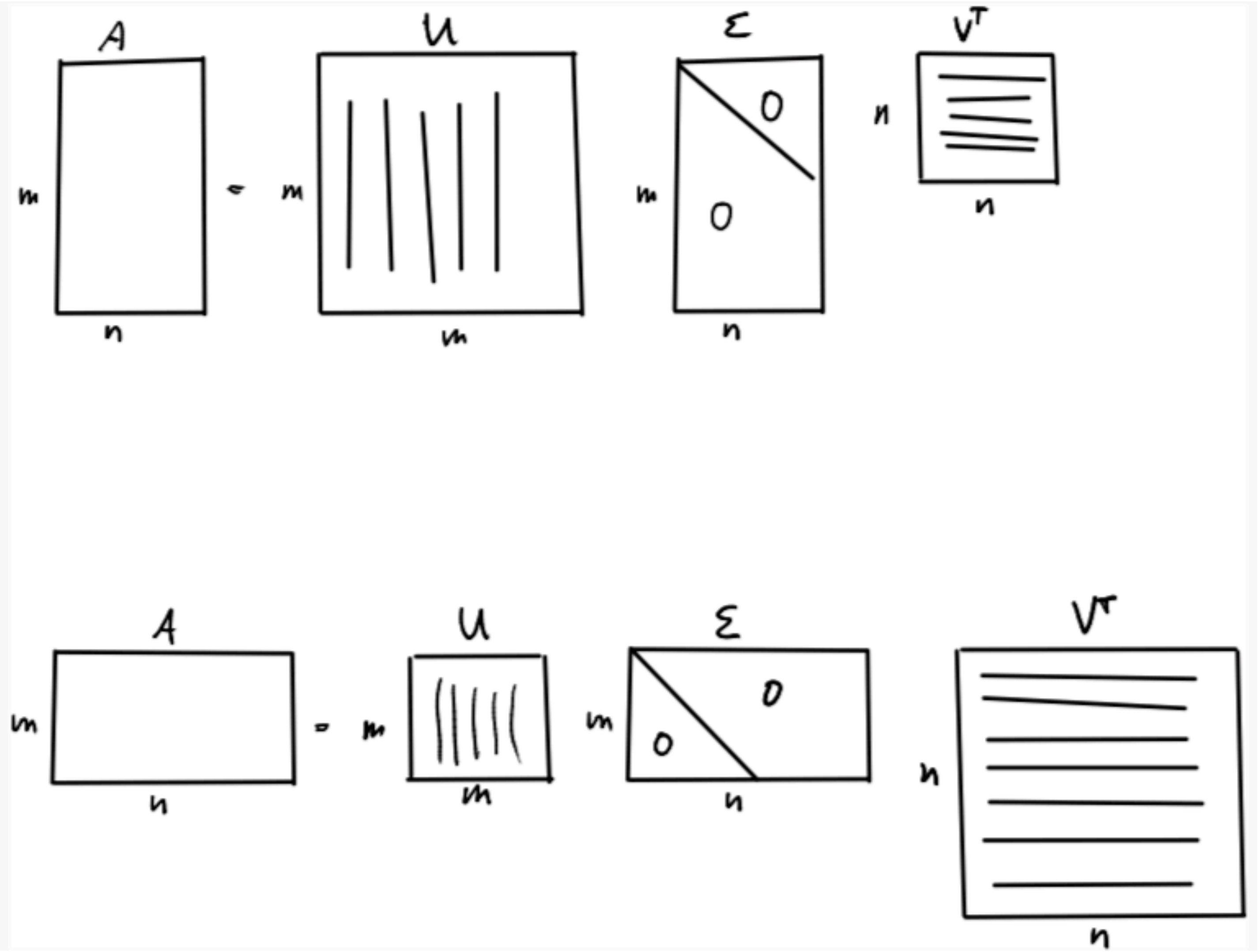
U is an m x m orthogonal matrix (Its columns are Left Singular Vectors)

Σ is an m x n diagonal matrix with **singular values** on the diagonal

$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & 0 \end{pmatrix} \qquad \text{Convention: } \sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \geqslant 0$$

V^T is an n x n orthogonal matrix (V's columns are called Right Singular Vectors) such that $\mathbf{A = U\Sigma V^T}$

Singular Value Decomposition



Evaluation of U and V

1. Find an orthogonal diagonalization of A^TA
   - Find the eigenvalues of A^TA and corresponding orthonormal set of eigenvectors
2. Set up V and Σ
   - Arrange the eigenvalues of A^TA in decreasing order and compute the square roots of the eigen values. Σ will be same size as A with D(diagonal entries are non zero singular values) in upper left corner and with 0's elsewhere
3. Derive U for $\mathbf{A = U\Sigma V^T}$

Evaluation of U and V - Example

Construct singular value decomposition of $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$

1.Find eigenvalues of $A^T A$

$$A^T A = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

Eigenvalues of $A^T A$ are $\lambda_1 = 360$, $\lambda_2 = 90$ and $\lambda_3 = 0$

2. Set up V and Σ

$$V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3] = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix} \rightarrow$$

corresponding

Unit eigen vectors

Evaluation of U and V - Example

3. Set up Σ

The square roots of the eigen values are singular values

$\sigma_1 = 6\sqrt{10}$, $\sigma_2 = 3\sqrt{10}$, $\sigma_3 = 0$

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix}, \qquad \Sigma = [\,D \quad 0\,] = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}$$

4.Construct U

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{6\sqrt{10}} \begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{3\sqrt{10}} \begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$$

Singular value decomposition of A is

$$A = \underset{\uparrow \atop U}{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}} \underset{\uparrow \atop \Sigma}{\begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}} \underset{\uparrow \atop V^T}{\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}}$$

Comparison between Eigenvalue decomposition and SVD

Eigenvalue Decomposition	Singular Value Decomposition
1. Works only for matrix that is always square	Works for rectangular matrix
2.Non diagonal matrices P and P^{-1} are inverses of each other	Non diagonal matrices U and V are not necessarily inverse of one another
3.Entries of D can be any complex number – negative, positive , imaginary	Entries in the diagonal matrix Σ are real and non negative, singular values are decreasing.
4. Vectors in eigenvalue decomposition matrix P are not necessarily orthogonal	Matrices U and V in SVD are orthonormal

•Left and Right Singular Vectors

- The column vectors in V are called right singular vectors
- The column vectors in U are called left singular vectors

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix}$$

$$U = \begin{bmatrix} -0.2425 & -0.9701 \\ -0.9701 & 0.2425 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.4472 & 0.8944 \\ -0.8944 & -0.4472 \end{bmatrix}$$

u_1, u_2 are left singular vectors

v_1, v_2 are right singular vectors

•Summation form of SVD

Let A be an m by n matrix of **rank r**. Let $A = \mathrm{U}\Sigma V^T$

Then A can be expanded as

$$A = \sigma_1(u_1 v_1^T) + \sigma_2(u_2 v_2^T) + \ldots\ldots + \sigma_r(u_r v_r^T)$$

Here $\sigma_1 \geq \sigma_2 \geq \ldots\ldots\ldots \geq \sigma_r$

It is addition of rank-1 matrices

•Summation Formula Example

Consider the following matrix

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix}$$

$$\text{Rref}(A) = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

Rank =1

• Eigendecomposition of AA^T

$$AA^T = \begin{bmatrix} 5 & 20 \\ 20 & 80 \end{bmatrix}$$

P1= $\begin{bmatrix} -0.2425 & -0.9701 \\ -0.9701 & 0.2425 \end{bmatrix}$

D1= $\begin{bmatrix} 85 & 0 \\ 0 & 0 \end{bmatrix}$

• Eigendecomposition of $A^T A$

$$A^T A = \begin{bmatrix} 17 & 34 \\ 34 & 68 \end{bmatrix}$$

P2= $\begin{bmatrix} 0.4472 & -0.8944 \\ 0.8944 & 0.4472 \end{bmatrix}$

D2= $\begin{bmatrix} 85 & 0 \\ 0 & 0 \end{bmatrix}$

•Summation Formula Example

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix}$$

Rank =1

$$U = \begin{bmatrix} -0.2425 & -0.9701 \\ -0.9701 & 0.2425 \end{bmatrix}$$

$$S = \begin{bmatrix} 9.2195 & 0 \\ 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.4472 & 0.8944 \\ -0.8944 & -0.4472 \end{bmatrix}$$

$$A = \sigma_1 u_1 v_1^T$$

$$A = 9.2195 \begin{bmatrix} -0.2425 \\ -0.9701 \end{bmatrix} \begin{bmatrix} -0.4472 & -0.8944 \end{bmatrix}$$

•Singular Values of A

It may be observed that the 2 singular values of A are the square root of eigenvalues of $\boldsymbol{AA^T}$ or $\boldsymbol{A^T A}$

- $\sigma_1 = \sqrt{\lambda_1}$ =9.2195 = $\sqrt{85}$
- $\sigma_2 = \sqrt{\lambda_2}$ =0 = $\sqrt{0}$

Face Recognition

Problem Statement : How to match a given image with a set of images in a database?

5 images each of

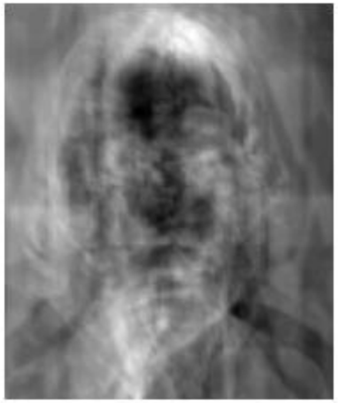
1. Abdul Kalaam
2. Kamal Hassan
3. Aishwarya Rai
4. Rahul Dravid
5. Virendra Sehwag

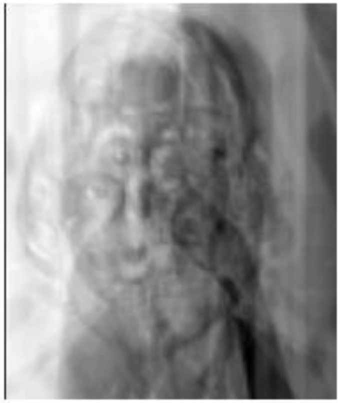
Comparing a new face with all these 25

Faces

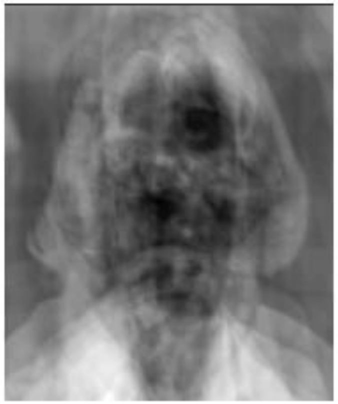
Average faces

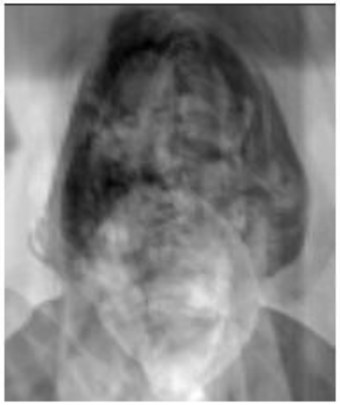
Eigen Faces





Singular Values





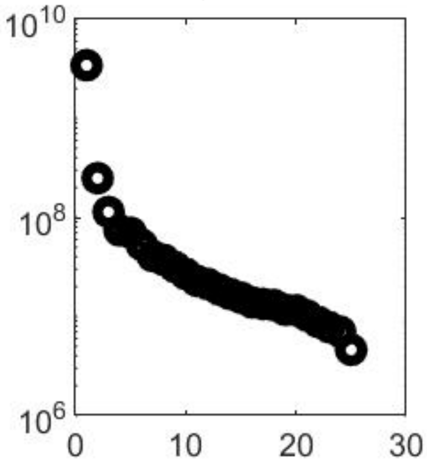
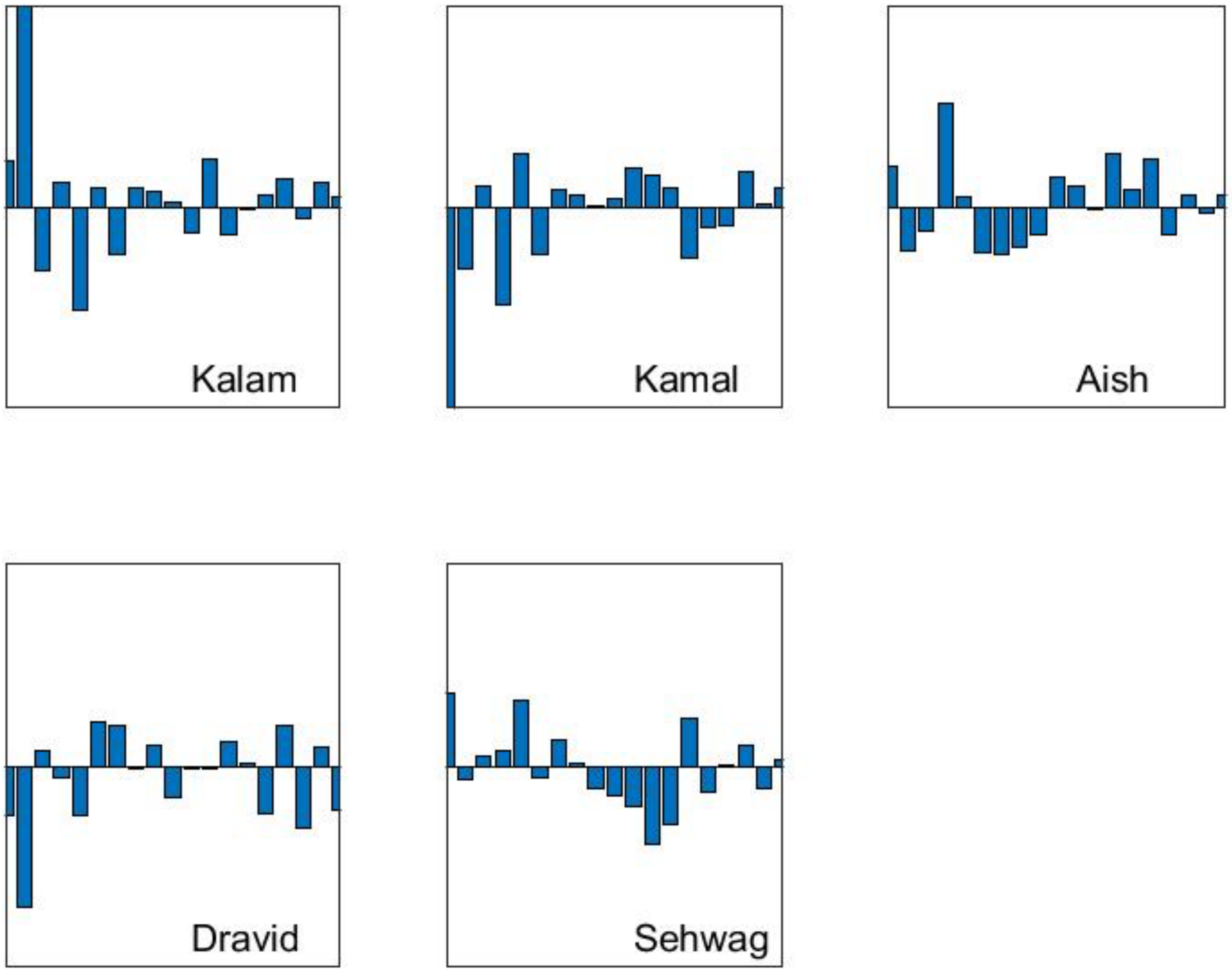
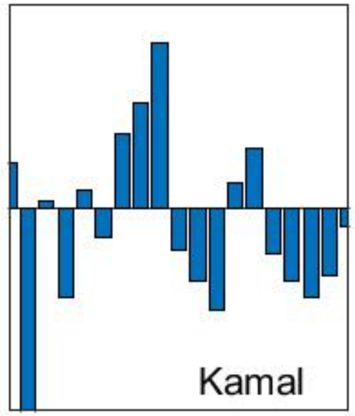


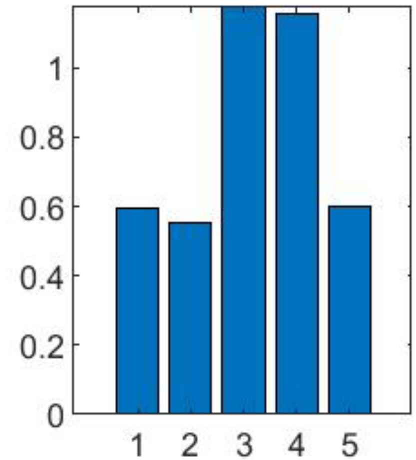
Image Keys



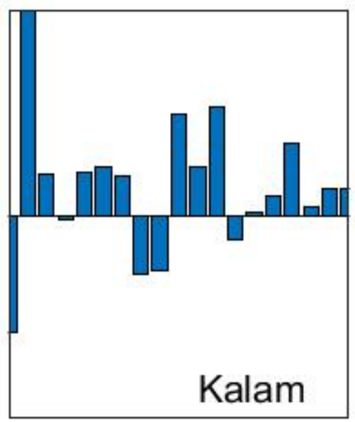
Face Comparison Errors

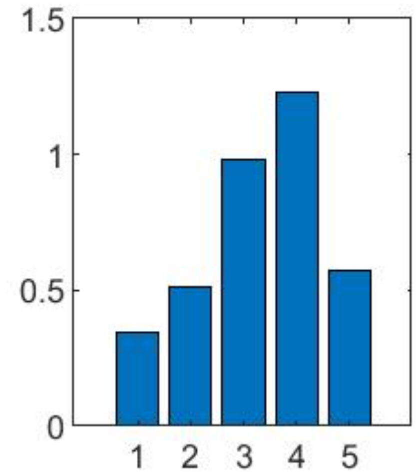
New Face





New Face





Principal Component Analysis and Dimensionality Reduction

σ_is dictate the dimension which are relevant

Suppose we consider only σ_1, σ_2 , σ_k $k < n$, then we have them as principal components and dimension is reduced to k from n