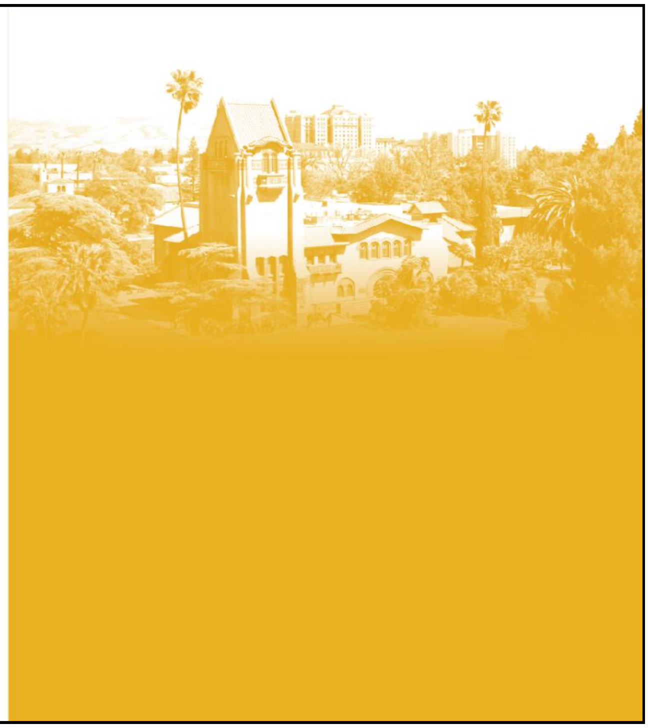
SJSU SAN JOSÉ STATE UNIVERSITY

Basic Linear Algebra



1

SJSU SAN JOSÉ STATE UNIVERSITY

Agenda

- Vectors & Matrices
- Linear System of Equations
- Eigenvalue Problem
- Linear Regression
- Singular Value Decomposition (SVD)
- Principal Component Analysis (PCA)

2

2

SJSU SAN JOSÉ STATE UNIVERSITY

Linear Algebra in Data Science

- A branch of mathematics that deals with linear equations and their representations as vectors and matrices.
- Linear algebra plays a crucial role in data science & machine learning algorithms.
- Understanding linear algebra allows us to efficiently perform computations on large datasets, manipulate images and audio signals, and solve optimization problems.
- Some common applications of linear algebra in data science include:
  - Linear Regression for fitting models to data
  - Singular Value Decomposition (SVD) for matrix factorization
  - Principal Component Analysis (PCA) for dimensionality reduction
  - Image processing and computer vision
  - Natural language processing and text analysis
  - Network analysis and graph theory

3

3

SJSU SAN JOSÉ STATE UNIVERSITY

Basic Concepts

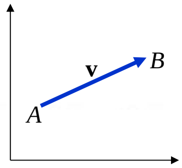
4

Vectors

- Numbers are called scalars.
- A vector is ordered list of numbers
  - numbers in the list are elements (entries/coefficients)

$$\begin{bmatrix} -1.1 \\ 0.0 \\ 3.6 \\ -7.2 \end{bmatrix} \text{ or } \begin{pmatrix} -1.1 \\ 0.0 \\ 3.6 \\ -7.2 \end{pmatrix}$$

- Vectors can be manipulated using Linear Algebra operations such as addition, multiplications and transformations
- A vector corresponds to a <u>displacement</u> from point A to point B.



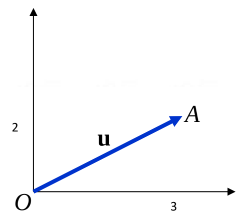
- You can represent a vector $\mathbf{v}$ from point A to point B, as a directed line segment from A to B, written $\overrightarrow{AB}$.

5

5

Vectors

- Every point A in the plane corresponds to a vector whose tail is at the origin O and the head is point A.
  - Example: If the coordinates of point A are (3, 2), we can write vector $\mathbf{u} = \overrightarrow{OA} = [3, 2]$.
- We can represent the vector $\mathbf{v}$ as a row vector $[3, 2]$ or as a column vector $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$

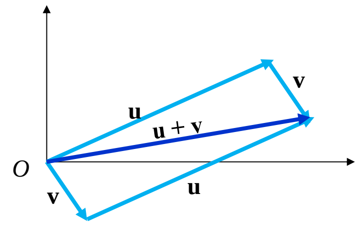


6

6

Basic Vector Operations

Suppose **u** = [3, 2] and **v** = [1, -1].



- Vector Addition: **u** + **v** = [3, 2] + [1, -1] = [3 + 1, 2 - 1] = [4, 1]
  - Add the corresponding components: 3 + 1 = 4 and 2 - 1 = 1
  - Note that **u** + **v** = **v** + **u**.
- Scalar multiplication: scale a vector by a constant
  - multiply each component of the vector by the constant.
  - Example: If **u** = [3, 2] , then 2**u** = [6, 4]

7

Vector-Vector Product

- Inner or Dot Product

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

- Outer Product

$$xy^T \in \mathbb{R}^{m\times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

9

Length or Norm of a Vector

The length (or norm) of a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ is the nonnegative scalar:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = \sqrt{\sum_{i=1}^{n} {v_i}^2}$$

The norm is also known as the Euclidean or L_2 norm.

10

Other Popular Norms

There are other popular norms as well

- L_1 or Manhattan norm:

$$\|\mathbf{v}\|_1 = \sum_{i=1}^{n} |v_i|$$

- L_∞ or Maximum or Chebyshev norm:

$$\|\mathbf{v}\|_\infty = \max_i |v_i|$$

11

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Norms of a Vector

Given $\boldsymbol{x} = [2, -3, 0.5]$ then

- L_1 or Manhattan norm:

$$\|x\|_1 = |2| + |-3| + |0.5| = 2 + 3 + 0.5 = 5.5$$

- L_2 or Euclidean norm:

$$\|x\|_2 = \sqrt{2^2 + (-3)^2 + 0.5^2} = \sqrt{4 + 9 + 0.25} = \sqrt{13.25} = 3.64$$

- L_∞ or Maximum or Chebyshev norm:

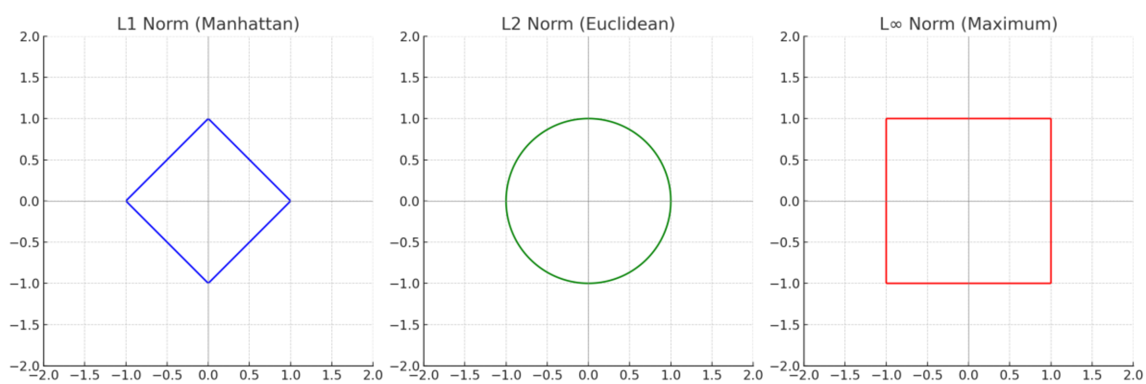
$$\|x\|_\infty = \max(|2|, |-3|, |0.5|) = 3$$

12

12

SJSU SAN JOSÉ STATE UNIVERSITY

Visual Comparison of Norms (2D)

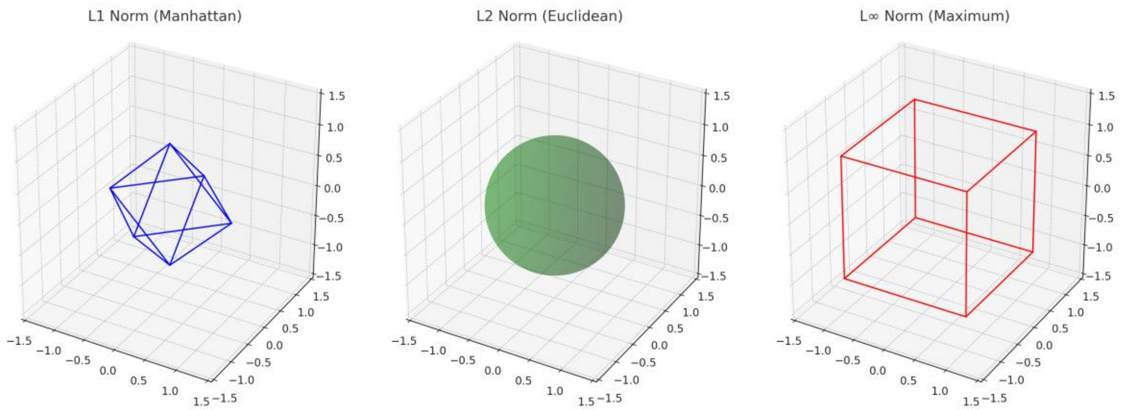


13

13

SJSU SAN JOSÉ STATE UNIVERSITY

Visual Comparison of Norms (3D)



SJSU SAN JOSÉ STATE UNIVERSITY

Normalization of a Vector

A unit vector is a vector of length 1.

– Examples: $\mathbf{e}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$ are unit vectors.

- If vector $\mathbf{v}$ is nonzero, we can find a unit vector in the same direction as $\mathbf{v}$ by dividing $\mathbf{v}$ by its own length.
- Finding a unit vector in the same direction as as vector $\mathbf{v}$ is called normalizing a vector.

Example: Vector Normalization

Example: If vector $\boldsymbol{v} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$ then $||\boldsymbol{v}|| = \sqrt{14}$.

The unit vector in the same direction as $\boldsymbol{v}$ is:

$$\boldsymbol{u} = \left(\frac{1}{||\mathbf{v}||}\right)\boldsymbol{v} = \frac{\mathbf{1}}{\sqrt{14}}\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{14} \\ -1/\sqrt{14} \\ 3/\sqrt{14} \end{bmatrix}$$

Linear Combination

A vector $\boldsymbol{v}$ is a linear combination of vectors $v_1, v_2, \ldots, v_k$ if there are scalars $c_1, c_2, \ldots, c_k$ such that $\boldsymbol{v} = c_1v_1 + c_2v_2 + \ldots + c_kv_k$.

Example: Vector $\begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}$ is a linear combination of vectors $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$, and $\begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix}$:

$$3\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + 2\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}$$

SJSU SAN JOSÉ STATE UNIVERSITY

Linear Independence

- A set of vectors are linearly independent if none of the vectors can be written as a linear combination of the others:
  - In other words, a set of vectors is linearly independent if no vector in the set is redundant and can be expressed in terms of the other vectors in the set
  - Two vectors are linearly independent if they point in different directions in space
  - A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is linearly independent if

$$a_1\boldsymbol{v}_1 + \ldots + a_n\boldsymbol{v}_n = \mathbf{0} \rightarrow a_1 = a_2 = \ldots = a_n = \mathbf{0}$$

- A set of vectors is linearly dependent if at least one of the vectors in the set can be expressed as a linear combination of the other vectors in the set
  - A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is linearly dependent if there exist scalar coefficients $a_1, a_2, \ldots,$ a_n not all zero such that $a_1\boldsymbol{v}_1 + \ldots + a_n\boldsymbol{v}_n = \mathbf{0}$

18

18

SJSU SAN JOSÉ STATE UNIVERSITY

Matrices

A matrix is a two-dimensional array of values (or a collection of vectors).

Examples:

$$A = \begin{bmatrix} \sqrt{5} & -1 & 0 \\ 2 & \pi & 1/2 \end{bmatrix} \qquad C = [7]$$

$$E = [1 \quad 12 \quad 5]$$

$$B = \begin{bmatrix} 5.1 & 1.2 & -1 \\ 6.9 & 0 & 4.4 \\ -7.3 & 9 & 8.5 \end{bmatrix} \qquad F = \begin{bmatrix} -6 \\ 14 \end{bmatrix}$$

19

19

SJSU SAN JOSÉ STATE UNIVERSITY

Special Matrices

- Zero matrix is a square matrix with all zeros

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- Identity matrix is a square matrix with ones on the diagonal and zeros everywhere else:

$$\boldsymbol{I}_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Diagonal matrix is a matrix where all non-diagonal elements are 0. This is typically denoted as D = diag(d_1, d_2, . . . , d_n)

$$\boldsymbol{D}_{ij} = \begin{cases} d_i & i = j \\ 0 & i \neq j \end{cases} \qquad \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$$

20

20

SJSU SAN JOSÉ STATE UNIVERSITY

Matrix-Vector Product

- If we write A by rows, then we can express Ax as:

$$y = Ax = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

- If we write A by columns, then we have:

$$y = Ax = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a^1 \end{bmatrix} x_1 + \begin{bmatrix} a^2 \end{bmatrix} x_2 + \ldots + \begin{bmatrix} a^n \end{bmatrix} x_n$$

linear combination of the columns of A

21

21

SJSU SAN JOSÉ STATE UNIVERSITY

Matrix-Matrix Multiplication

- Matrices can be viewed as a collection of vectors.
- Matrix-matrix multiplication is a set of vector-vector or dot products.

$$C = AB = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b^1 & b^2 & \cdots & b^p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b^1 & a_1^T b^2 & \cdots & a_1^T b^p \\ a_2^T b^1 & a_2^T b^2 & \cdots & a_2^T b^p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b^1 & a_m^T b^2 & \cdots & a_m^T b^p \end{bmatrix}$$

- Or as a set of matrix-vector products:

$$C = AB = A \begin{bmatrix} | & | & & | \\ b^1 & b^2 & \cdots & b^n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab^1 & Ab^2 & \cdots & Ab^n \\ | & | & & | \end{bmatrix}$$

SJSU SAN JOSÉ STATE UNIVERSITY

Matrix-Matrix Multiplication

- Associative: $(AB)C = A(BC)$
- Distributive: $A(B + C) = AB + BC$
- In general, matrix multiplication is not commutative: $AB \neq BA$
- For example, if $A \in R^{m \times n}$ and $B \in R^{n \times p}$, the matrix product BA does not even exist if m and p are not equal!)

SJSU SAN JOSÉ STATE UNIVERSITY

Matrix Properties

- Transpose: $(\boldsymbol{A}^T)_{ij} = \boldsymbol{A}_{ji}$

$$(\boldsymbol{A}^T)^T = \boldsymbol{A}$$
$$(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$$
$$(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$$

- Symmetric: $\boldsymbol{A} = \boldsymbol{A}^T$

- Trace: $\mathrm{tr}(\boldsymbol{A}) = \sum_i \boldsymbol{A}_{ii}$

$$\mathrm{tr}(\boldsymbol{A}^T) = \mathrm{tr}(\boldsymbol{A})$$
$$\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$$
$$\mathrm{tr}(b\boldsymbol{A}) = \mathrm{b}\ \mathrm{tr}(\boldsymbol{A})$$
$$\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$$

24

24

SJSU SAN JOSÉ STATE UNIVERSITY

Rank of a Matrix

- The column rank of a matrix $\boldsymbol{A} \in R^{m\times n}$ is the largest number of columns of $\boldsymbol{A}$ that constitute a linearly independent set.
- The row rank of a matrix $\boldsymbol{A} \in R^{m\times n}$ is the largest number of rows of $\boldsymbol{A}$ that constitute a linearly independent set.
- It turns out that the column rank of $\boldsymbol{A}$ = the row rank of $\boldsymbol{A}$ so both quantities are referred to collectively as the rank of A, denoted as rank(A).
- For $A \in R^{m\times n}$, $rank(A) \leq \min(m,n)$. If $rank(A) = \min(m,n)$, $\boldsymbol{A}$ is said to be full rank.

25

25

SJSU SAN JOSÉ STATE UNIVERSITY

Inverse of a Square Matrix

- The inverse of a square matrix $A \in R^{n \times n}$ is denoted A^{-1} and is the unique matrix such that

$$A^{-1}A = AA^{-1} = I$$

- A is invertible or non-singular if A^{-1} exists and non-invertible or singular otherwise.
- In order for a square matrix A to have an inverse A^{-1}, then A must be full rank.
- Properties (Assuming $A, B \in R^{n \times n}$ are non-singular):
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^T = (A^T)^{-1}$

SJSU SAN JOSÉ STATE UNIVERSITY

Orthogonal Matrices

- Two vectors $\boldsymbol{x}, \boldsymbol{y} \in R^n$ are orthogonal if $\boldsymbol{x}^T\boldsymbol{y} = 0$.
- A vector $\boldsymbol{x} \in R^n$ is normalized if $\|\boldsymbol{x}\|_2 = 1$.
- A square matrix $\boldsymbol{U} \in R^{n \times n}$ is orthogonal if all its columns are orthogonal to each other and are normalized (the columns are then referred to as being orthonormal).
- Properties:
  - The inverse of an orthogonal matrix is its transpose ($\boldsymbol{U}^{-1} = \boldsymbol{U}^T$).

$$\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^T = I$$

  - Operating on a vector with an orthogonal matrix will not change its Euclidean norm, i.e.,

$$\|\mathbf{U}\,\mathbf{v}\|_2 = \|\mathbf{v}\|_2$$

for any $\boldsymbol{x} \in R^n$, $\boldsymbol{A} \in R^{n \times n}$ orthogonal.

SJSU SAN JOSÉ STATE UNIVERSITY

The Determinant

- The determinant of a square matrix $\boldsymbol{A} \in R^{n\times n}$, is a function $\det : R^{n\times n} \rightarrow R$, and is denoted |A| or det A.
- Given a matrix $\begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_n^T & — \end{bmatrix}$

consider the set of points $S \subset R^n$ as follows:

$$S = \{v \in \mathbb{R}^n : v = \sum_{i=1}^{n} \alpha_i a_i \text{ where } 0 \le \alpha_i \le 1, i = 1, \ldots, n\}$$

- The absolute value of the determinant of A is a measure of the "volume" of the set S.

SJSU SAN JOSÉ STATE UNIVERSITY

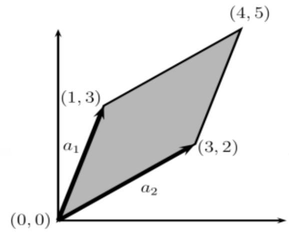
Visualization of The Determinant

- For example, consider the 2 × 2 matrix,

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$$

- Here, the rows of the matrix are

$$a_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad a_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$



SJSU SAN JOSÉ STATE UNIVERSITY

Properties of The Determinant

- For $\boldsymbol{A} \in R^{n\times n}$, $|\boldsymbol{A}| = |\boldsymbol{A}^T|$.
- For $\boldsymbol{A}, \boldsymbol{B} \in R^{n\times n}$, $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$.
- For $\boldsymbol{A} \in R^{n\times n}$, $|\boldsymbol{A}| = 0$ if and only if $\boldsymbol{A}$ is singular (i.e., non-invertible).
- For $\boldsymbol{A} \in R^{n\times n}$ and $\boldsymbol{A}$ is non-singular, $|\boldsymbol{A}^{-1}| = 1/|\boldsymbol{A}|$.

30

30

SJSU SAN JOSÉ STATE UNIVERSITY

Linear System of Equations

31

SJSU SAN JOSÉ STATE UNIVERSITY

Matrix Equations

- N unknowns (x_1, x_2, …, x_N) and N equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N &= b_2 \\ &\vdots \\ a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{NN}x_N &= b_N \end{aligned}$$

- These equations have unique solution if all equations are linearly-independent.
- Matrix form: $[\mathbf{A}]\cdot\{\mathbf{x}\} = \{\mathbf{b}\}$

$$[\mathbf{A}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix} \quad \{\mathbf{x}\} = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{Bmatrix}, \quad \{\mathbf{b}\} = \begin{Bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{Bmatrix}$$

SJSU SAN JOSÉ STATE UNIVERSITY

Matrix Equations

- Solution to the linear system: $[\mathbf{A}]^{-1}$ exists or $[\mathbf{A}]$ is not singular

$$\begin{aligned} [\mathbf{A}]^{-1}[\mathbf{A}]\cdot\{\mathbf{x}\} &= [\mathbf{A}]^{-1}\cdot\{\mathbf{b}\} \\ [\mathbf{I}]\cdot\{\mathbf{x}\} &= [\mathbf{A}]^{-1}\cdot\{\mathbf{b}\} \\ \{\mathbf{x}\} &= [\mathbf{A}]^{-1}\cdot\{\mathbf{b}\} \end{aligned}$$

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Linear System of Equations

Solve the following linear system:

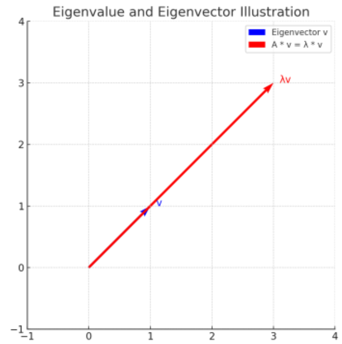
$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$$

SJSU SAN JOSÉ STATE UNIVERSITY

Eigenvalue Problem
&
Eigendecomposition

SJSU SAN JOSÉ STATE UNIVERSITY

Eigenvalues and Eigenvectors



- Eigenvalue problem: $[\mathbf{A}]\cdot\{\mathbf{x}\} = \lambda\{\mathbf{x}\}$
  - λ : Eigenvalues
  - $\{\mathbf{x}\}$: Eigenvectors
- How to solve?

$$[\mathbf{A}]\cdot\{\mathbf{x}\} - \lambda\{\mathbf{x}\} = \{\mathbf{0}\}$$
$$[\mathbf{A}]\cdot\{\mathbf{x}\} - \lambda[\mathbf{I}]\{\mathbf{x}\} = \{\mathbf{0}\}$$
$$[\mathbf{A} - \lambda\mathbf{I}]\cdot\{\mathbf{x}\} = \{\mathbf{0}\}$$

  - In order to have non-trivial solution, the determinant must be zero:

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

  - Calculate λ from this equation and calculate $\{\mathbf{x}\}$ from the eigenvalue problem.

36

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Eigenvalue Problem

Find the eigenvalues and eigenvectors of the matrix: $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

37

37

SJSU SAN JOSÉ STATE UNIVERSITY

What is Eigendecomposition?

- Eigendecomposition is the process of breaking a square matrix into its eigenvalues and eigenvectors.
- If a matrix $\boldsymbol{A} \in R^{n \times n}$ has $\boldsymbol{n}$ linearly independent eigenvectors, it can be written as:

$$\boldsymbol{A} = \boldsymbol{\Phi} \, \boldsymbol{\Lambda} \, \boldsymbol{\Phi}^{-1}$$

- Eigendecomposition is possible when $\boldsymbol{A}$ is diagonalizable, i.e., it has $\boldsymbol{n}$ linearly independent eigenvectors.
- All symmetric matrices (i.e., $\boldsymbol{A} = \boldsymbol{A}^T$) are always diagonalizable with real eigenvalues.

38

SJSU SAN JOSÉ STATE UNIVERSITY

Example:

- Perform eigendecompostion on $\boldsymbol{A} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$

39

Applications of Eigendecomposition

- Matrix powers:

$$A^k = \Phi\, \Lambda^{\mathrm{k}}\, \Phi^{-1}$$

  Efficient for computing powers of $\boldsymbol{A}$.

- Principal Component Analysis (PCA): Uses eigendecomposition of covariance matrices to reduce dimensionality.
- Differential equations: Solutions often involve eigenvalues and eigenvectors of a system matrix.
- Simplifies understanding of matrix behavior: scaling along directions defined by eigenvectors.

SJSU SAN JOSÉ STATE UNIVERSITY

Singular Value Decomposition

SJSU SAN JOSÉ STATE UNIVERSITY

Singular Value Decomposition

- A given matrix A can be decomposed as:

$$A_{m\times n} = U_{m\times m} S_{m\times n} V^*_{n\times n}$$

where U and V are unitary (orthogonal), and S is "diagonal"

$$A = USV^* = [u_1\ u_2\ \dots\ u_m] \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} v_1^* \\ v_2^* \\ \vdots \\ v_n^* \end{bmatrix}$$

$$S = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \text{ or } S = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \dots & 0 \\ 0 & 0 & \dots & \sigma_m & 0 & \dots & 0 \end{bmatrix}$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ real & positive

42

42

SJSU SAN JOSÉ STATE UNIVERSITY

Singular Value Decomposition

- The columns of U and V are called the left and right singular vectors

$$U = [u_1\ u_2\ \dots\ u_m]$$
$$V = [v_1\ v_2\ \dots\ v_n].$$

$$A = USV^* \Rightarrow AV = US \Rightarrow Av_k = \sigma_k u_k$$
$$\Rightarrow A^*U = VS^* \Rightarrow A^* u_k = \sigma_k v_k$$

$$\left.\begin{matrix} Av_k = \sigma_k u_k \\ A^* u_k = \sigma_k v_k \end{matrix}\right\} A^*Av_k = \sigma_k^2 v_k$$

eigenvalues & eigenvectors of A*A

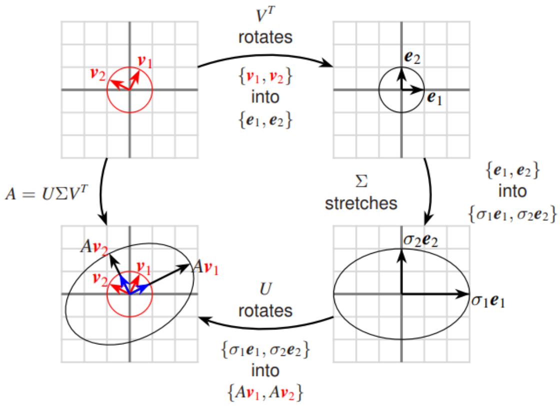
43

43

SJSU SAN JOSÉ STATE UNIVERSITY

Singular Value Decomposition Illustration

- SVD = rotation + scaling + rotation



SJSU SAN JOSÉ STATE UNIVERSITY

SVD Truncation

- Recall

$$A = USV^* = [u_1 \; u_2 \; \dots \; u_m] \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} v_1^* \\ v_2^* \\ \vdots \\ v_n^* \end{bmatrix} \Rightarrow A = [\sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \dots + \sigma_n u_n v_n^*]$$

- We can form a “truncated” version of A with fewer # of terms:

$$A_k = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \dots + \sigma_k u_k v_k^*$$

- Error of truncation: $A_n - A_k = \sigma_{k+1} u_{k+1} v_{k+1}^* + \dots + \sigma_n u_n v_n^*$

SVD Example

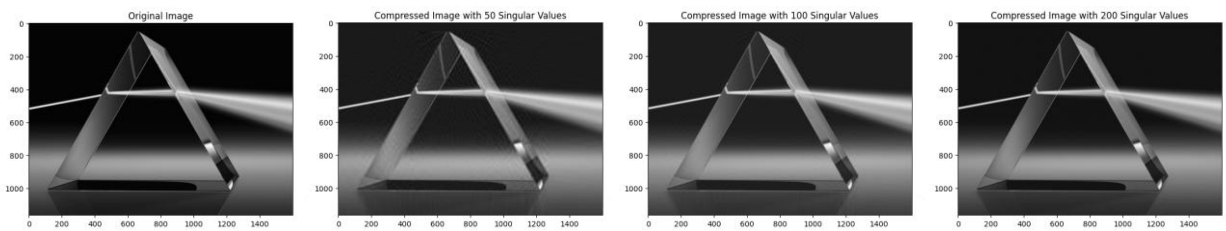
- Consider the 4×5 matrix

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{U} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{5} & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{V}^* = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ -\sqrt{0.2} & 0 & 0 & 0 & -\sqrt{0.8} \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix}$$

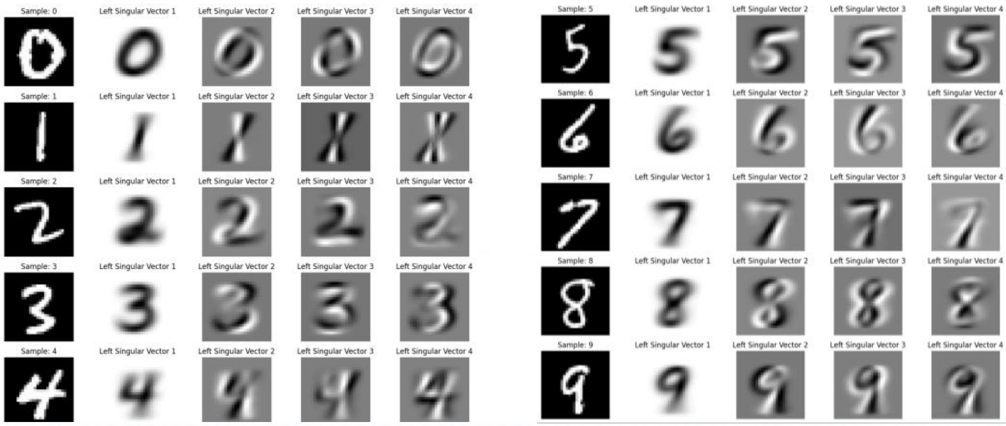
SVD Example: Image Processing

- SVD can be applied to image compression:



SJSU SAN JOSÉ STATE UNIVERSITY

MNIST Dataset with SVD

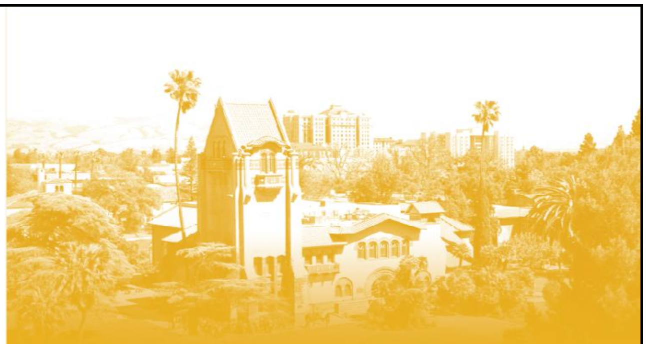


48

48

SJSU SAN JOSÉ STATE UNIVERSITY

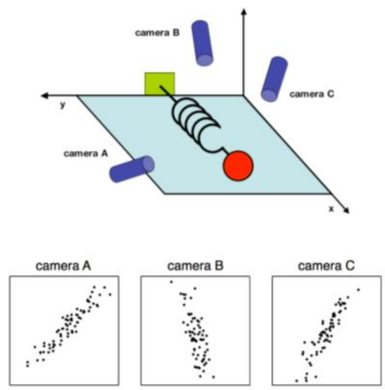
Principal Component Analysis



49

Principal Component Analysis (PCA)

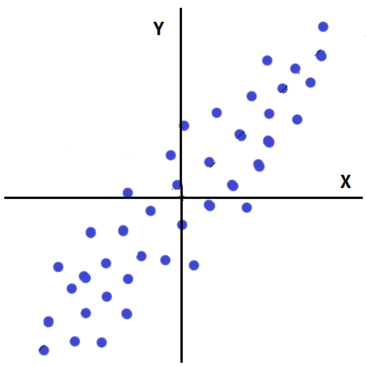
- PCA: A numerical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components
- The original data are projected onto a much smaller space, resulting in dimensionality reduction
- Method: Find the eigenvectors of the covariance matrix, and these eigenvectors define the new space

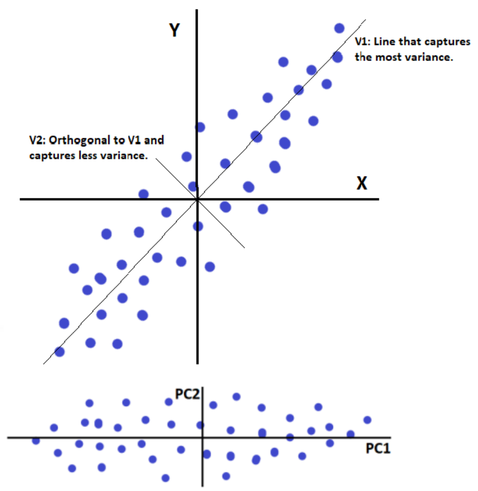


50

Graphical Explanation of PCA

- Consider a data set with the following scatter plot:



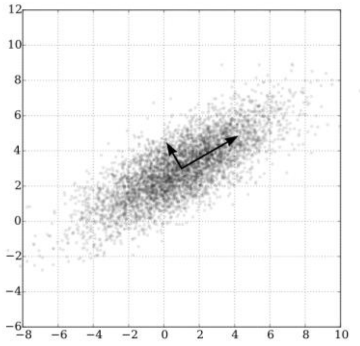


51

SJSU SAN JOSÉ STATE UNIVERSITY

Principal Component Analysis (Method)

- PCA Steps
  - Standardize the Data (i.e. zero mean & unit standard deviation)
  - Compute covariance matrix
  - Compute eigenvalues & eigenvectors
  - Sort eigenvalues
  - Select principal components (top k eigenvectors)
  - Transform the Data
- Works for numeric data only



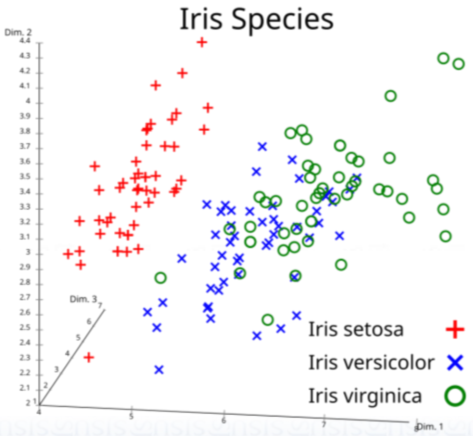
52

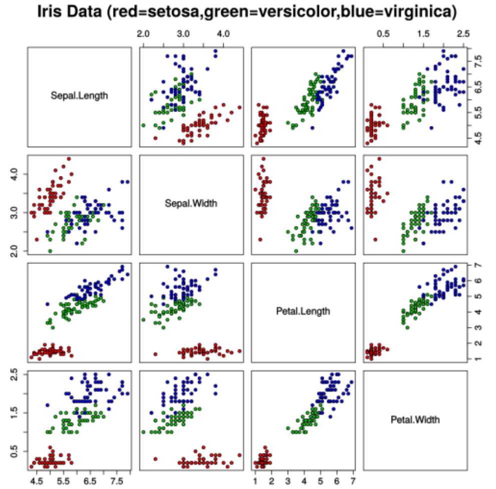
52

SJSU SAN JOSÉ STATE UNIVERSITY

PCA Example: Iris Dataset

- Iris flower dataset has 4 features, 3 targets:





53

53

SJSU SAN JOSÉ STATE UNIVERSITY

PCA Example: Iris Dataset