



BITS Pilani
Pilani Campus

Applied Machine Learning

Dr. Harikrishnan N B
Computer Science and Information Systems

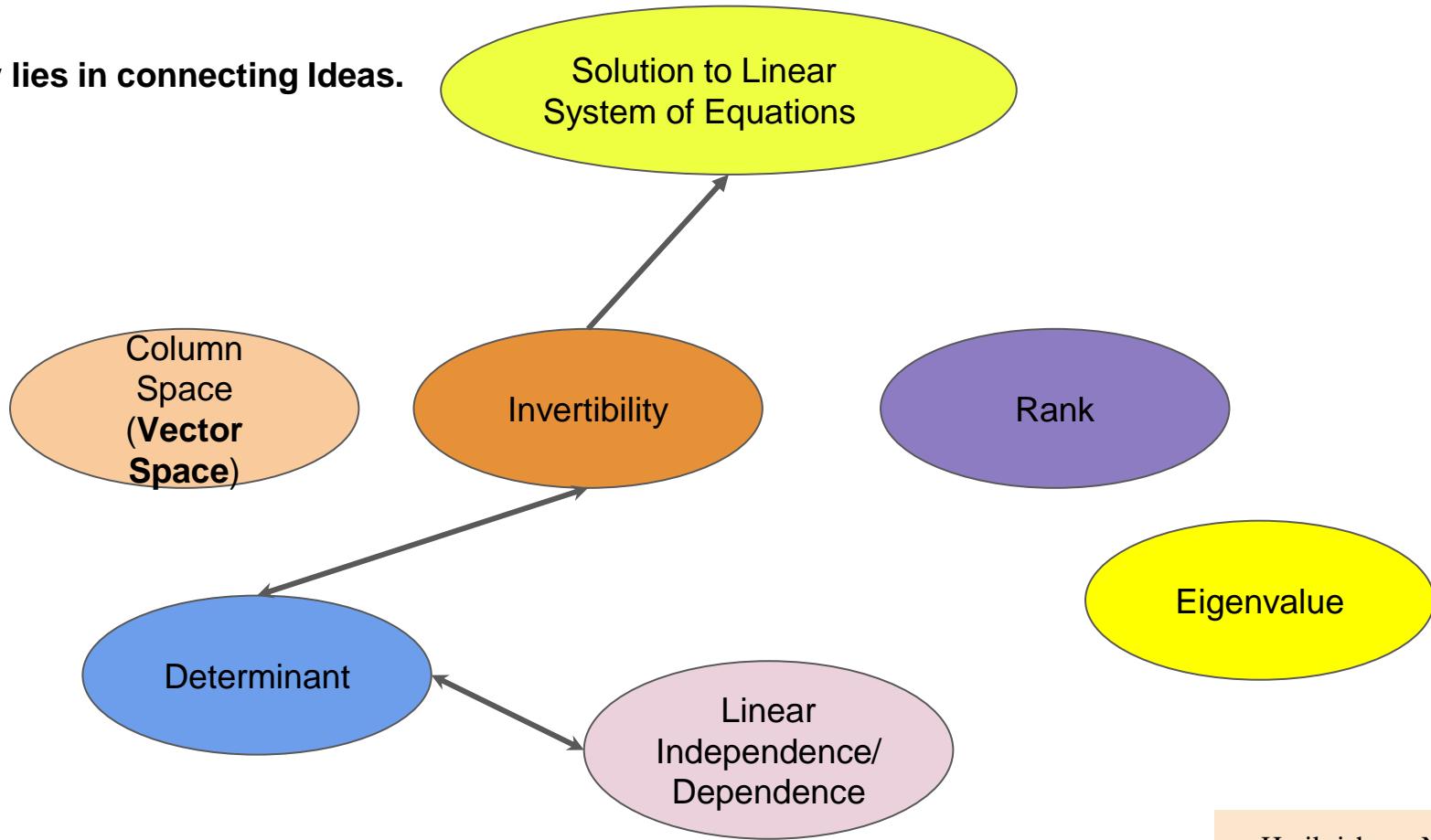


SE ZG568 / SS ZG568, Applied Machine Learning Lecture No. 5 [16- Feb-2025]

Recap

Basics of Linear Algebra,
Row Picture, Col Picture, Algebraic Way
Solution to System of Linear Equations
Inverse of a Matrix
Linear Regression,
PCA

Beauty lies in connecting Ideas.



Orthogonal and Orthonormal Matrix

Orthogonal vectors

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0$$

$$\text{L2 - norm} = \sqrt{2}$$

Orthonormal vectors

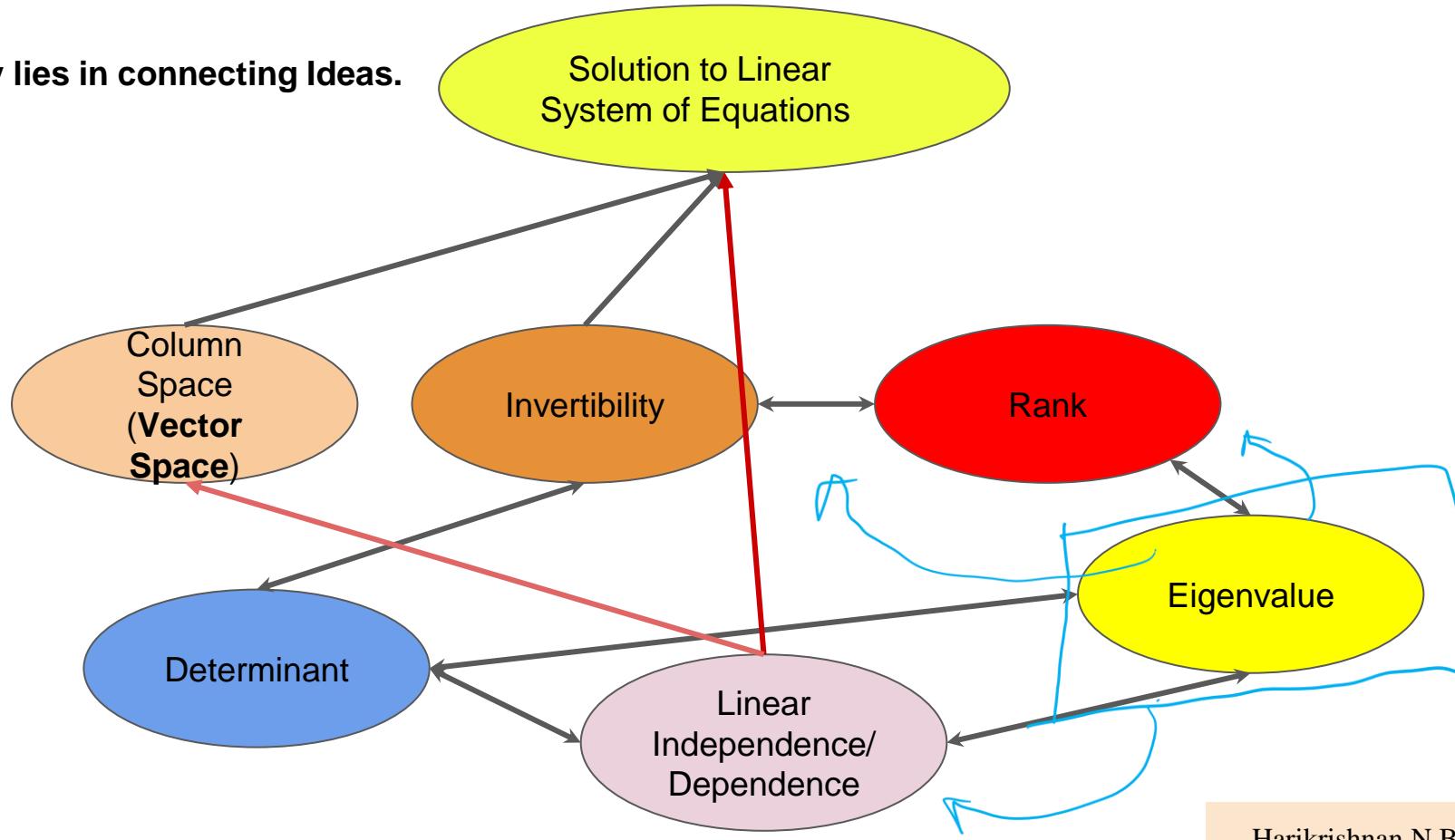
$$\begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = 0$$

$$\text{L2 - norm} = 1$$

Orthonormal Matrix

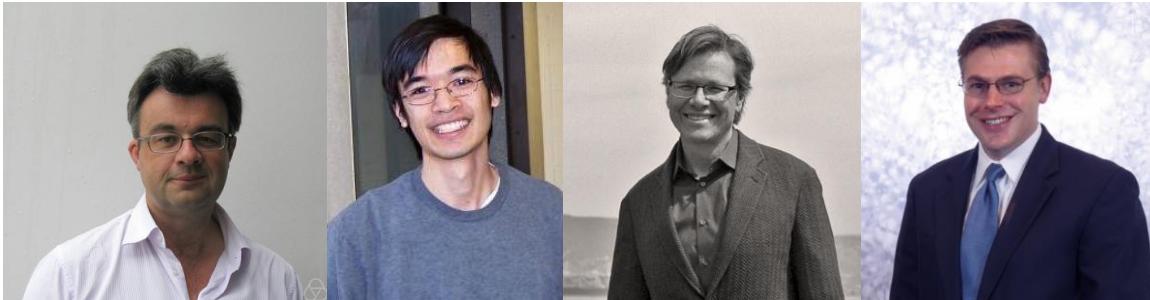
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad A^{-1} = \frac{1}{|A|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Beauty lies in connecting Ideas.



Do we know everything about $y = Ax$?

[Emmanuel Candès](#) [Terence Tao](#) [David Donoho](#) [Justin Romberg](#)



Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?

Emmanuel Candes[†] and Terence Tao[#]

[†] Applied and Computational Mathematics, Caltech, Pasadena, CA 91125

[#] Department of Mathematics, University of California, Los Angeles, CA 90095

**An Introduction To
Compressive Sampling**

A sensing/sampling paradigm that goes against the common knowledge in data acquisition

Emmanuel J. Candès and Michael B. Wakin

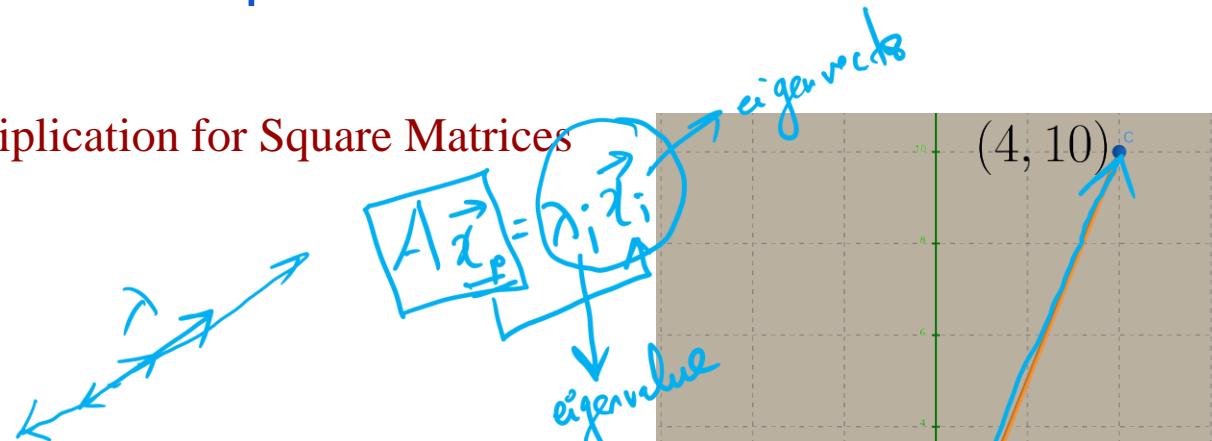
Conventional approaches to sampling signals or images follow Shannon's celebrated theorem: the sampling rate must be at least twice the maximum frequency present in the signal (the so-called Nyquist rate). In fact, this principle underlies nearly all signal acquisition protocols used in consumer audio and visual electronics, medical imaging devices, radio receivers, and so on. (For some signals, such as images that are not naturally bandlimited, the sampling rate is dictated not by the Shannon theorem but by the desired temporal or spatial resolution. However, it is common in such systems to use an antialiasing low-pass filter to bandlimit the signal before sampling, and so the Shannon theorem plays an implicit role.) In the field of data conversion, for example, standard analog-to-digital converter (ADC) technology implements the usual quantized Shannon representation: the signal is uniformly sampled at or above the Nyquist rate.

Digital Object Identifier 10.1109/NSP.2007.914721

Matrix Vector Multiplication as a Transformation

Intuition for Matrix vector multiplication for Square Matrices

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 10 \end{bmatrix}$$

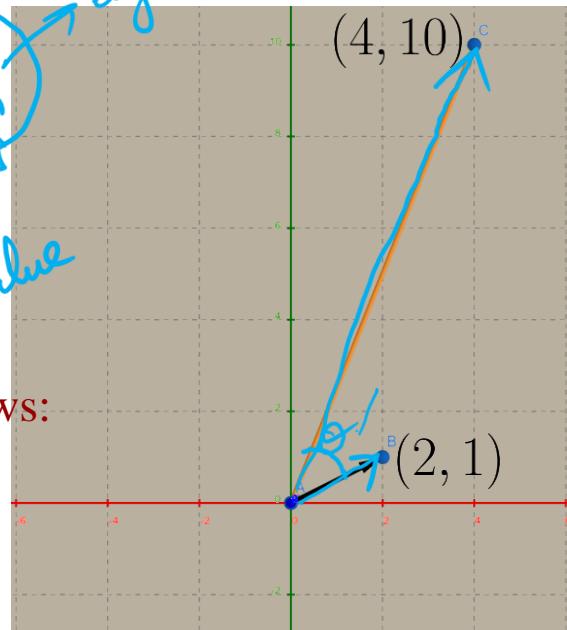


Matrix(Square Matrix) vector multiplication can be seen as follows:

- Rotation ✓
- Stretching or Shrinking ✓

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 2y \end{bmatrix}$$

$$\begin{bmatrix} & \\ & \end{bmatrix} \alpha f^2$$



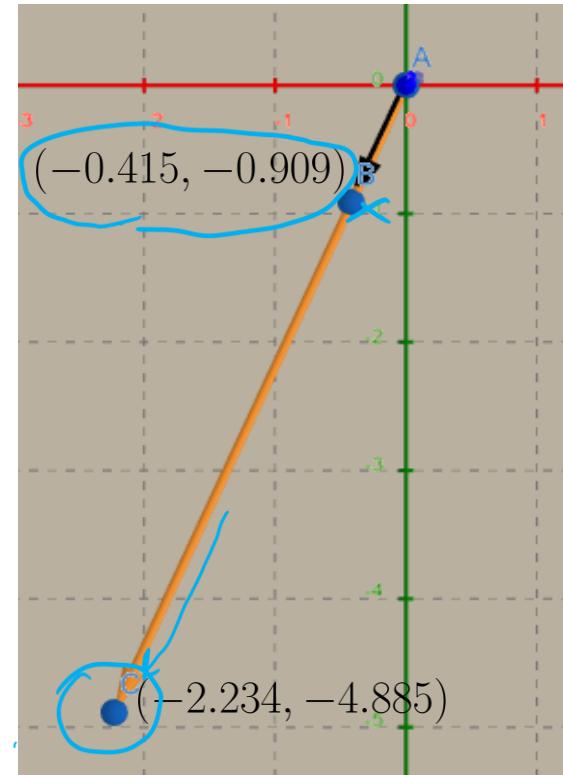
Special Vectors

$$\begin{array}{c}
 A \\
 \boxed{\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}} \quad \vec{x} \\
 \downarrow \quad \downarrow \\
 \begin{bmatrix} -0.415 \\ -0.909 \end{bmatrix} = \begin{bmatrix} -2.234 \\ -4.885 \end{bmatrix} = 5.372 \begin{bmatrix} -0.415 \\ -0.909 \end{bmatrix} \\
 \downarrow \quad \downarrow \\
 A\vec{x} = \lambda\vec{x}
 \end{array}$$

eigenvalue λ

$$A\vec{x} = \lambda\vec{x}$$

1. Direction of \vec{x} is unchanged. (No rotation)
2. Only the magnitude is scaled by a factor λ
3. \vec{x} - **eigenvector of matrix A**
4. λ - **eigenvalue of matrix A**



$$A = \begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix}$$

Suppose \vec{x} is an eigenvector of 'A'
Size of \vec{x} 2×1 $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, λ is

$$A\vec{x} \Rightarrow$$

$$\left[\begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right] = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$A\vec{x} = \lambda\vec{x}$$

$$A\vec{x} - \lambda\vec{x} = \vec{0}$$

$$\left[\begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right] - \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$A\vec{x} = \lambda\vec{x}$$

$$A\vec{x} - \lambda\vec{x} = \vec{0}$$

$$\Leftrightarrow \begin{bmatrix} \vec{x} \\ 2 \times 1 \end{bmatrix} (\begin{bmatrix} A - \lambda I \\ 2 \times 2 \end{bmatrix}) = \vec{0} \times$$

$$(A - \lambda I) \vec{x} = \vec{0}$$

$$\left(\begin{bmatrix} A \\ 2 \times 2 \end{bmatrix} - \begin{bmatrix} \lambda \cdot 1 & 0 \\ 0 & \lambda \cdot 1 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

$$A\vec{x} - \begin{bmatrix} \lambda \cdot 1 & 0 \\ 0 & \lambda \cdot 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

$$A\vec{x} - \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \end{bmatrix} = \vec{0}$$

$$A\vec{x} - \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

$$3 \cdot 2 + 4 \cdot 2 \Rightarrow 2 \cdot (3+4)$$

$$(A - \lambda I) \vec{x} = \text{scale} = \text{Brut} \cdot 2$$

$$A = \begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix}$$

$$A\vec{x} = \lambda\vec{x}$$

$$(A - \lambda I) \cdot \vec{x} = 0$$

$$(A - \lambda I) \cdot \vec{x} = 0$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$\begin{bmatrix} -a_1^T & - \\ -a_2^T & - \end{bmatrix} \begin{bmatrix} \vec{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$A\vec{x} = \lambda\vec{x}$

$a_1^T \vec{x} = 0$, $a_2^T \vec{x} = 0$

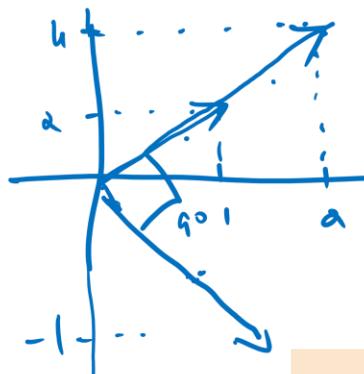
non zero vector

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \vec{x} = ?$$

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 1 \times \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

$$col_2 \Rightarrow 2 \cdot col_1$$



$$(A - \lambda I) \cdot \vec{z} = \vec{0}$$

Linearly Dependent

$$\det(A - \lambda I) = 0$$

$$A = \begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix}$$

$$\det(A - \lambda I) = 0$$

$$\begin{bmatrix} 5-\lambda & 3 \\ 0 & 2-\lambda \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix} \xrightarrow{\text{det}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow \det(A) = 0$$

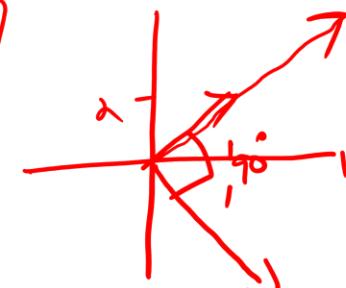
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Can you find a non zero $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ which satisfies the above equation?

$$\left\{ \begin{array}{l} \\ \end{array} \right. \quad \rightarrow$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Linearly Dependent



$$\boxed{\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = 0}$$

$$(5-\lambda)(2-\lambda) - 3 \cdot 0 = 0$$

$$(5-\lambda)(2-\lambda) - 3 \cdot 0 = 0$$

$$\lambda = 5 \text{ or } \lambda = 2$$

$\lambda = 5$, I want to find eigenvectors

$$(\vec{A} - \lambda \vec{I}) \cdot \vec{x} = \vec{0}$$

$$\boxed{\lambda = 5}$$

$$\begin{bmatrix} 0 & 3 \\ 0 & -3 \end{bmatrix} \cdot \vec{x} = \vec{0}$$

$$\left| \begin{array}{l} \begin{bmatrix} 0 & 3 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{array} \right.$$

for $\lambda = 2$, I want to find eigenvectors

$$(\vec{A} - \lambda \vec{I}) \cdot \vec{x} = \vec{0}$$

$$\vec{A} \vec{x}_1 = \lambda_1 \vec{x}_1$$

$$\vec{A} \vec{x}_2 = \lambda_2 \vec{x}_2$$

$\lambda = 2$

$$\begin{bmatrix} 3 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\vec{A} \vec{x}_1 = \begin{pmatrix} 5 \\ 0 \end{pmatrix} = 5 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\vec{A} \vec{x}_2 = \begin{pmatrix} 2 \\ -2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Eigenvalues and Eigenvectors

- For an $n \times n$ square matrix A , there are ' n ' eigenvalues and ' n ' eigenvectors. Let $\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n$ be the ' n ' eigenvectors and $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ are the corresponding eigenvalues.

$$\boxed{A\vec{x}_1 = \lambda_1\vec{x}_1}$$
$$\boxed{A\vec{x}_2 = \lambda_2\vec{x}_2}$$
$$\boxed{A\vec{x}_3 = \lambda_3\vec{x}_3}$$
$$\cdot$$
$$\boxed{A\vec{x}_n = \lambda_n\vec{x}_n}$$

$$X = \begin{bmatrix} & & & & \\ \vec{x}_1 & \vec{x}_2 & \vec{x}_3 & \cdots & \vec{x}_n \\ & & & & \end{bmatrix}$$

$$A = \begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix} \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

\downarrow

$$X = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$$

$$A \underset{2 \times 2}{\times} X \underset{2 \times 2}{=} (\) \underset{2 \times 2}{=}$$

\vec{x}, \vec{y} are eigenvecs
of $A(n \times n)$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\lambda_1 = 5$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\lambda_2 = 2$$

$$\begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\begin{bmatrix} 5 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} = \begin{bmatrix} 5x_1 + 3x_2 & 5y_1 + 3y_2 \\ 0x_1 + 2x_2 & 0y_1 + 2y_2 \end{bmatrix}$$

$$AX = \begin{bmatrix} A\vec{x} \\ A\vec{y} \end{bmatrix} = \begin{bmatrix} \lambda_1 \vec{x} \\ \lambda_2 \vec{y} \end{bmatrix} = \begin{bmatrix} \vec{x} & \vec{y} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

A is a square matrix of size 2×2 . Let \vec{x}_1 and \vec{x}_2 be the eigenvectors of A with eigenvalues λ_1 and λ_2 respectively ($\lambda_1 \neq \lambda_2 \neq 0$)

$$X = \begin{bmatrix} 1 & 1 \\ \vec{x}_1 & \vec{x}_2 \\ 1 & 1 \end{bmatrix} \text{ stacking eigenvectors of } \underline{\underline{A}} \text{ as cols}$$

$$\underline{\underline{AX}} = \begin{bmatrix} 1 & 1 \\ A\vec{x}_1 & A\vec{x}_2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \lambda_1\vec{x}_1 & \lambda_2\vec{x}_2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \vec{x}_1 & \vec{x}_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$$\underline{\underline{A}} \begin{bmatrix} 1 & 1 \\ \vec{x}_1 & \vec{x}_2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \vec{x}_1 & \vec{x}_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$$AX = X\Lambda$$

$$A \underline{\underline{XX}} = X\Lambda X$$

$$A = X\Lambda X^{-1}$$

Very Very Important Part

- For an $n \times n$ square matrix A , there are ‘ n ’ eigenvalues and ‘ n ’ eigenvectors. Let $\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n$ be the ‘ n ’ eigenvectors and $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ are the corresponding eigenvalues.

$$A\vec{X} = \begin{bmatrix} A & | & \vec{x}_1 & \vec{x}_2 & \vec{x}_3 & \cdots & \vec{x}_n \end{bmatrix}$$

Spectral Decomposition

$$A \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vec{x}_3 \\ \vdots \\ \vec{x}_n \end{bmatrix} = \begin{bmatrix} A\vec{x}_1 \\ A\vec{x}_2 \\ A\vec{x}_3 \\ \vdots \\ A\vec{x}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \vec{x}_1 \\ \lambda_2 \vec{x}_2 \\ \lambda_3 \vec{x}_3 \\ \vdots \\ \lambda_n \vec{x}_n \end{bmatrix}$$

A blue bracket under the first matrix indicates that the columns are linearly independent.

$$\begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vec{x}_3 \\ \vdots \\ \vec{x}_n \end{bmatrix} = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vec{x}_3 \\ \vdots \\ \vec{x}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_n & & \end{bmatrix}$$

Spectral Decomposition

$$A = \begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ A & & & & \\ & & & & \\ & & & & \end{bmatrix} \begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ \vec{x}_1 & \vec{x}_2 & \vec{x}_3 & \cdots & \vec{x}_n \\ & & & & \\ & & & & \end{bmatrix} = \begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ \vec{x}_1 & \vec{x}_2 & \vec{x}_3 & \cdots & \vec{x}_n \\ & & & & \\ & & & & \end{bmatrix} \begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_n & & \end{bmatrix}$$

$$AX = X\Lambda$$

$$AXX^{-1} = X\Lambda X^{-1}$$

$$AI = X\Lambda X^{-1}$$

$$A = X\Lambda X^{-1}$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 9 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

$$AA^{-1} \text{ & } AAT$$

Practical Challenges and Important Points

When can we apply $A = X\Lambda X^{-1}$?

- A should be a square matrix ✓
- When A has 'n' linearly independent eigenvectors, then X^{-1} always exist.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}_{2 \times 2} \quad A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

What happens when A is Symmetric ($A^T = A$)?

- The eigenvectors of a symmetric matrix A can be chosen as ORTHONORMAL. So in this case X is orthonormal.
- For an ORTHONORMAL matrix X, the inverse is its transpose $X^{-1} = X^T$

$$A = X\Lambda X^{-1}$$

$$A = X\Lambda X^T$$

$$\begin{bmatrix} Y_{r_1} & Y_{r_2} \\ -Y_{r_2} & Y_{r_2} \end{bmatrix} A = \begin{bmatrix} W_L & Y_{r_2} \\ Y_{r_2} & -Y_{r_2} \end{bmatrix} \quad \tilde{A}^T$$

- The eigenvectors of a symmetric matrix A can be chosen as ORTHONORMAL. So in this case X is orthonormal. Why?

$$x_1 = \begin{pmatrix} \sqrt{r_1} \\ -\sqrt{r_1} \end{pmatrix}$$

$$X = \begin{bmatrix} | & | & \dots & | \\ x_1 & x_2 & \dots & x_n \\ | & | & \dots & | \end{bmatrix}$$

Rule 1 $\rightarrow \vec{x}_i \cdot \vec{x}_j = 0$ for all $i \neq j$

Rule 2 $\rightarrow \|\vec{x}_i\|_2 = 1$ for $i = 1$ to N



$$X^T X \Rightarrow$$

$$\begin{bmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \vec{x}_3^T \\ \vdots \\ \vec{x}_n^T \end{bmatrix}$$

$$\begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vec{x}_3 \\ \vdots \\ \vec{x}_n \end{bmatrix}$$

$$= \begin{bmatrix} \vec{x}_1^T \vec{x}_1 & \vec{x}_1^T \vec{x}_2 & \dots & \vec{x}_1^T \vec{x}_n \\ \vec{x}_2^T \vec{x}_1 & \vec{x}_2^T \vec{x}_2 & \dots & \vec{x}_2^T \vec{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{x}_n^T \vec{x}_1 & \vec{x}_n^T \vec{x}_2 & \dots & \vec{x}_n^T \vec{x}_n \end{bmatrix}$$



The eigenvectors of a symmetric matrix can be chosen orthonormal

$A^T = A$, let x_i, x_j be the eigenvectors of A and their corresponding eigenvalues be λ_i and λ_j . Note that $\lambda_i \neq \lambda_j$

$$\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$
$$\|x_1\|_2 = \sqrt{5} \quad \|x_2\|_2 = \sqrt{5}$$
$$Ax_1 = \lambda_1 x_1 \quad Ax_2 = \lambda_2 x_2$$
$$A^T x_1 = \lambda_1 x_1 \quad A^T x_2 = \lambda_2 x_2$$

$$A x_j = \lambda_j x_j$$

$$x_j^T A = (A^T x_j)^T$$
$$\begin{matrix} BC \\ \text{m} \times n \times m \\ (BC)^T \\ C^T B^T \end{matrix}$$

$$(BC)^T = C^T B^T$$
$$(A^T z_j)^T = z_j^T (A^T)^T = z_j^T A$$
$$(A^T z_j)^T z_i = z_j^T A z_i = \lambda_i z_j^T z_i = 0$$
$$z_j^T x_j^T x_i - \lambda_i z_j^T x_j^T x_i = 0$$
$$(z_j - \lambda_i)(z_j^T x_i) = 0$$

$$Ax_i = \lambda_i x_i$$
$$z_j^T A x_i = \lambda_i z_j^T x_i$$
$$(A^T z_j)^T z_i = \lambda_i z_j^T z_i$$
$$\lambda_j z_j^T z_i = \lambda_i z_j^T z_i$$

$$K \cdot Z$$
$$m \times r \times n$$

$$(KZ)^T$$

$$Z^T K^T$$

$$V_{rs} \quad a_{rs}$$

Practical Challenges and Important Points

When can we apply $A = X\Lambda X^{-1}$?

- A should be a square matrix
- When A has ‘n’ linearly independent eigenvectors, then X^{-1} always exist.

What happens when A is Symmetric ($A^T = A$)?

- The eigenvectors of a symmetric matrix A can be chosen as **ORTHONORMAL**. So in this case **X** is orthonormal.
- For an **ORTHONORMAL** matrix **X**, the inverse is its transpose $X^{-1} = X^T$

$$A = X\Lambda X^{-1}$$

$$A = X\Lambda X^T$$

Practical Challenges

What if A is not a square matrix?

- We cannot apply Spectral Decomposition.

Don't Worry!!!



Singular Value Decomposition works for any Matrix.

A Few more steps to PCA

What all minimum can we say about this data?

X	1	2	3	4	5
Y	1	5	4	6	7

X, Y are the features

What all minimum can we say about this data?

$$\text{Mean}(X) = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\text{Variance}(X) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2$$

$$\text{Cov}(X,Y) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

X	Y	var(X)	var(Y)	cov(X,Y)	cov(Y,X)
1	1				
2	5				
3	4				
4	6				
5	7				
Mean (X)	Mean(Y)	var(X)	var(Y)	cov(X,Y)	cov(Y,X)

$$\text{Mean}(X) = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\text{Variance}(X) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2$$

$$\text{Cov}(X, Y) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

X	Y
1	1
2	5
3	4
4	6
5	7

$$\text{Mean}(X) = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\text{Variance}(X) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2$$

$$\text{Cov}(X, Y) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

X	Y
1	1
2	5
3	4
4	6
5	7

What all minimum can we say about this data?

$$\text{Mean}(X) = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\text{Variance}(X) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2$$

$$\text{Cov}(X,Y) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

X	Y	var(X)	var(Y)	cov(X,Y)	cov(Y,X)
1	1				
2	5				
3	4				
4	6				
5	7				
Mean (X)	Mean(Y)	var(X)	var(Y)	cov(X,Y)	cov(Y,X)
3.0	4.6	2.5	5.3	2.25	2.25

Variance- Covariance Matrix

$$\begin{matrix} & X & Y \\ X & \begin{bmatrix} var(X) & cov(X, Y) \\ cov(Y, X) & var(Y) \end{bmatrix} \\ Y & \end{matrix}$$

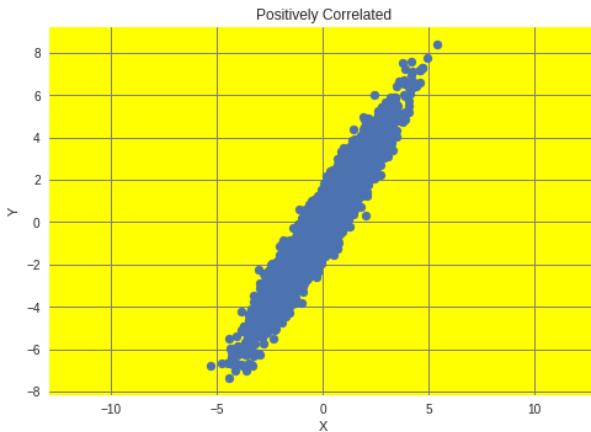
Recall the properties of a symmetric matrix!!!

- Variance - Covariance Matrix is symmetric. $cov(X, Y) = cov(Y, X)$
- The diagonal entries represents variance
- The off- diagonal entries represents the correlation of X and Y

What does Variance - Covariance Matrix signifies?

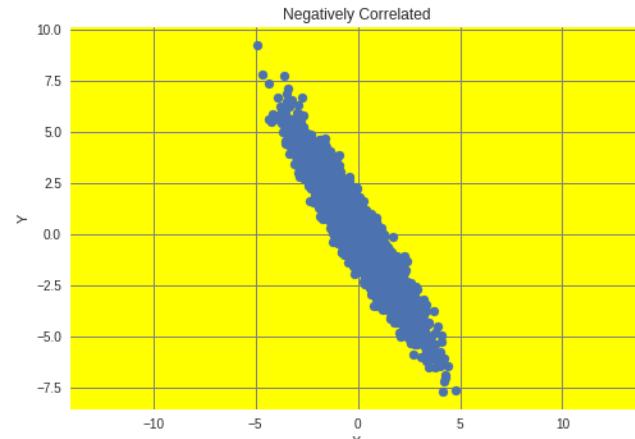
Case I

$$\begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$



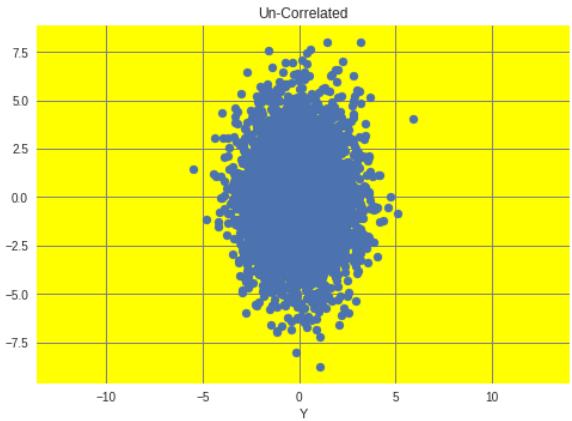
Case II

$$\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$$



Case III

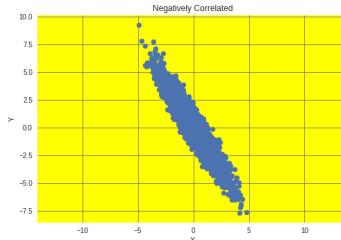
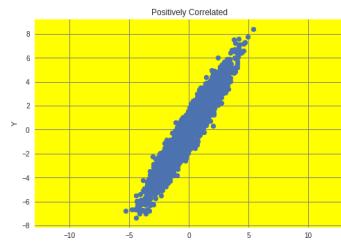
$$\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$$



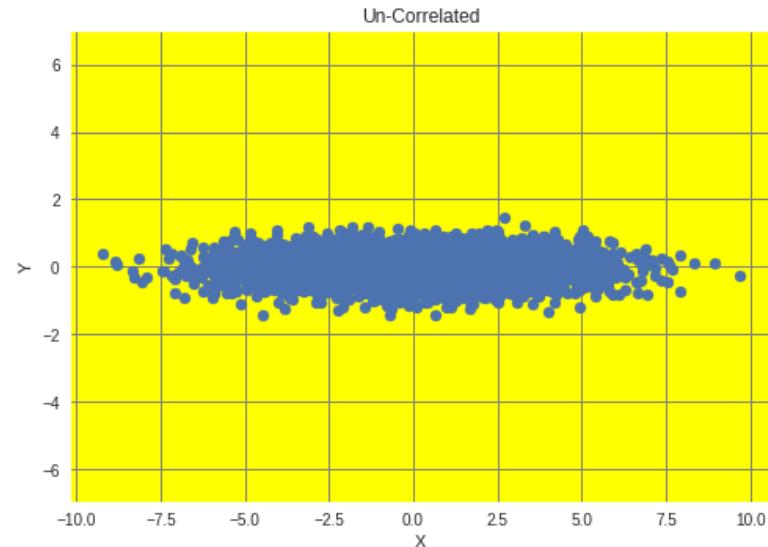
Note: In all cases mean is (0,0)

So what does PCA do ?

- Principal Component Analysis (PCA) makes the data **UNCORRELATED**.



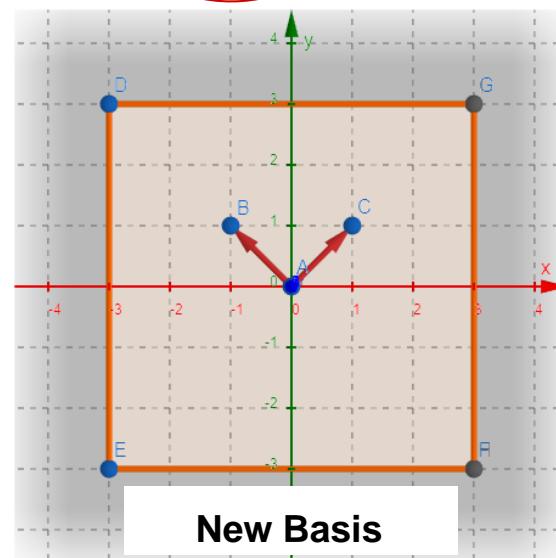
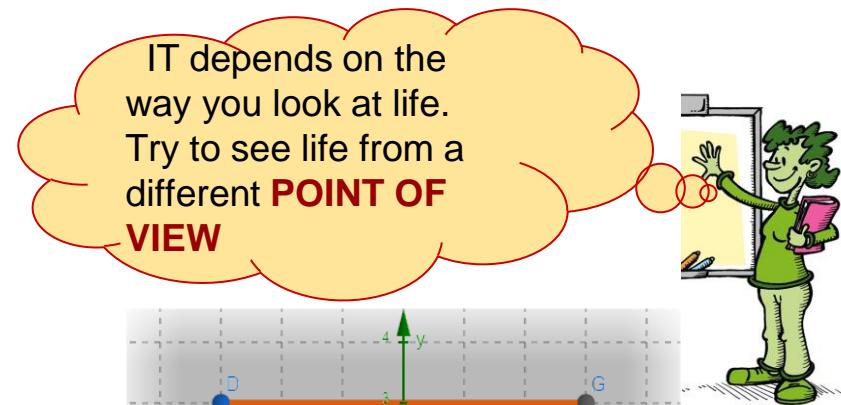
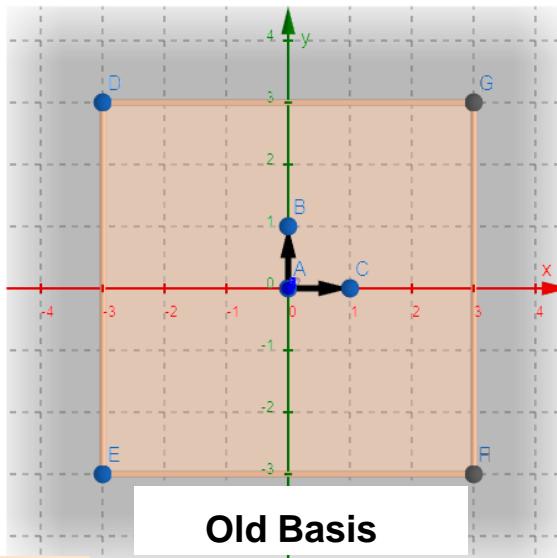
PCA achieves this by
Change of Basis



Change of Basis



My life is full of struggle.

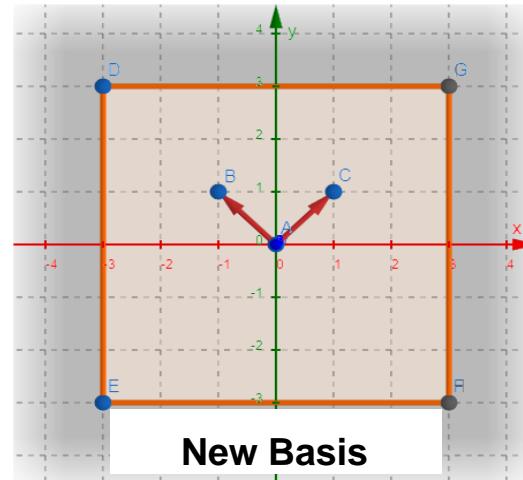
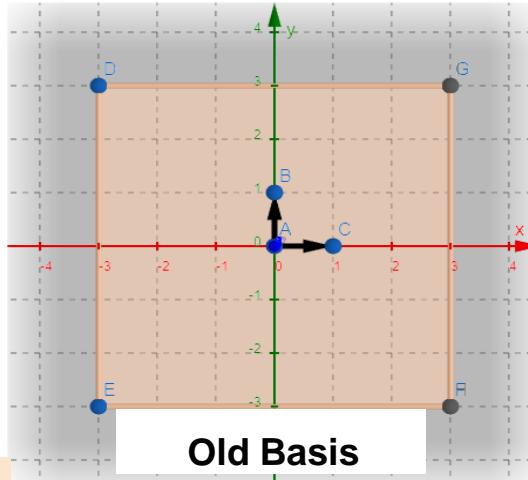


Recall

Dimension of a Vector space - Every vector space has a dimension. Dimension is the number of basis vectors required to span the vector space.

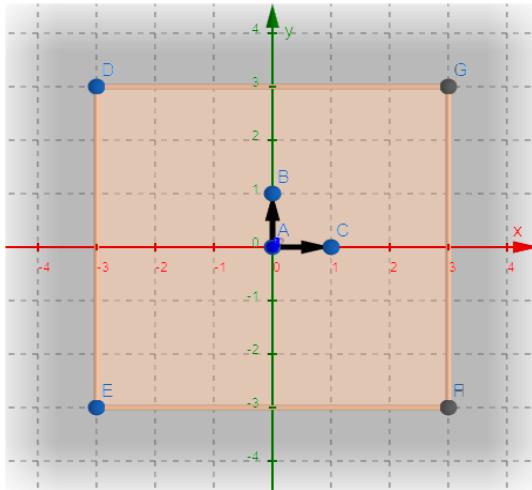
Properties of Basis Vectors -

- Basis vectors have to be linearly independent.
- Basis vectors should span the vector space.



Example of Change of Basis

To represent a point (2,3) in old basis and new basis- How to understand this?

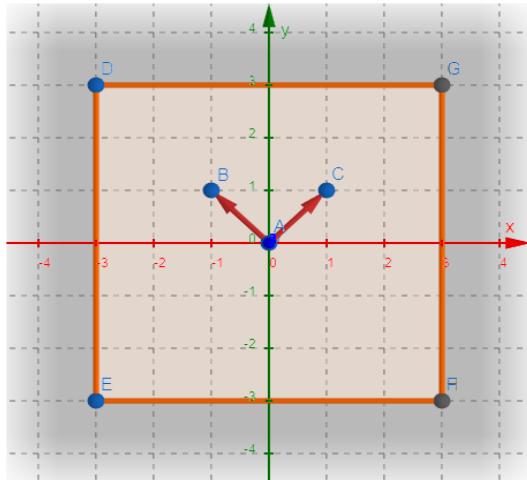


$$\text{Old basis} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

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

$$\begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

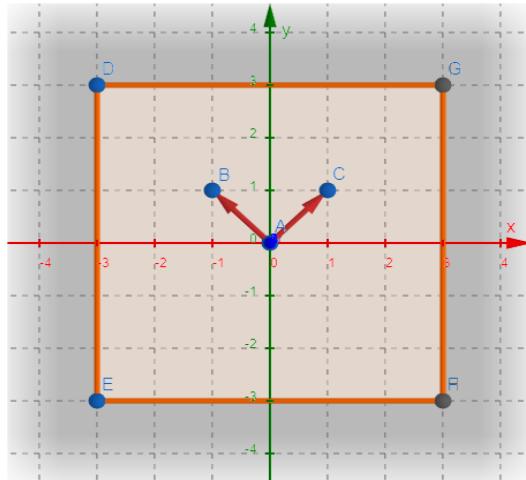
New Basis Representation



$$\text{New Basis} = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$$

$$x \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + y \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

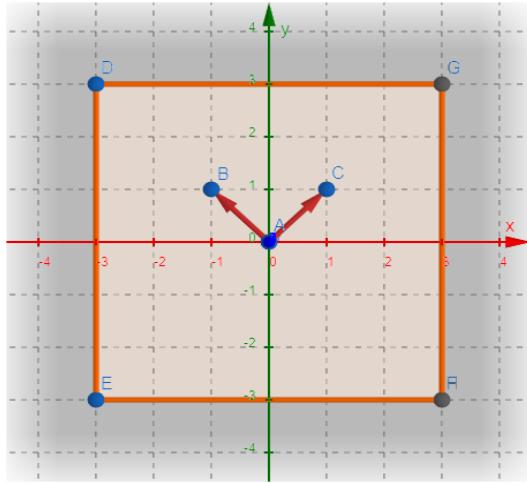
Finding x and y for representing (2,3) using new basis



$$x \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + y \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$



$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$



$$P\vec{x} = \vec{y}$$

$$P^{-1}P\vec{x} = P^{-1}\vec{y}$$

$$\vec{x} = P^{-1}\vec{y}$$

For **ORTHONORMAL MATRIX, $P^{-1} = P^T$**

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

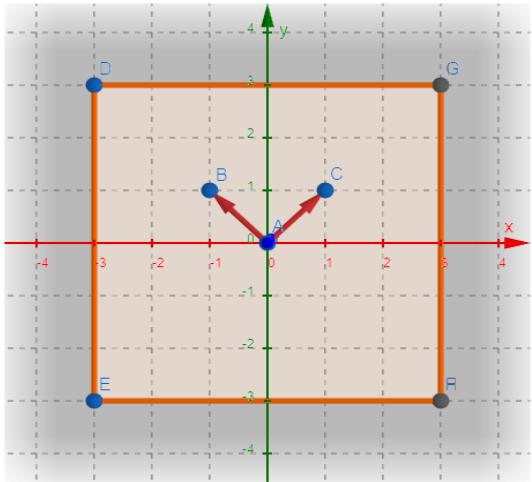
$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

In our case the matrix P is ORTHONORMAL

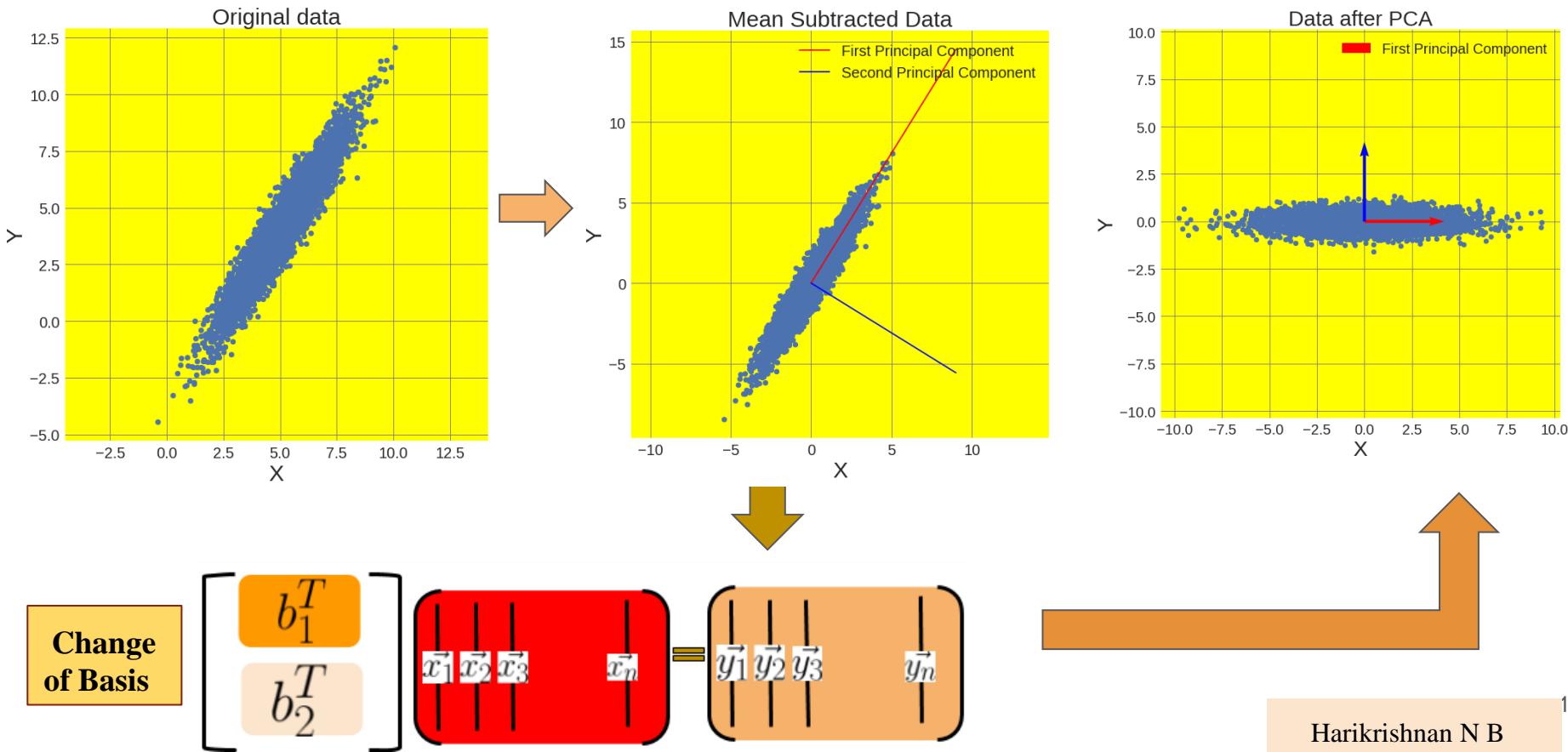
$$\vec{x} = P^{-1}\vec{y} = P^T\vec{y}$$

$$\begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{5}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\frac{5}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$



Steps in PCA

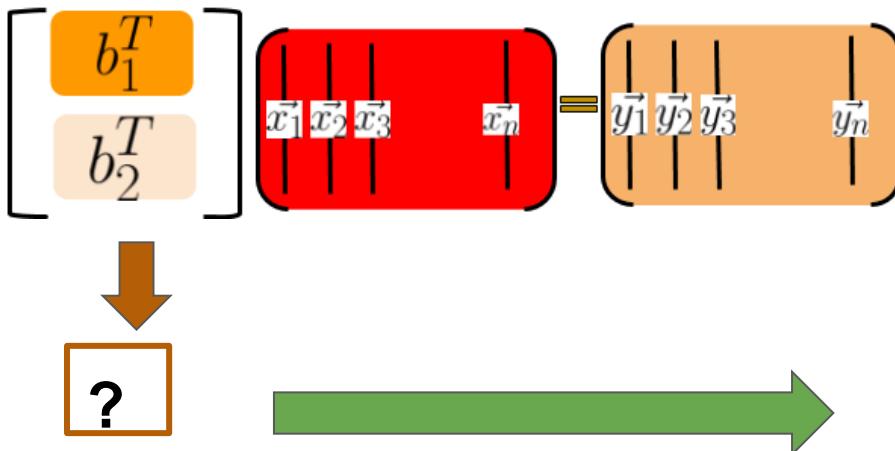


**Change
of Basis**

126

Harikrishnan N B

What should be the NEW BASIS so that DATA is UNCORRELATED?



Rows of matrix P are the **eigenvectors** of the **variance - covariance matrix** of the **mean subtracted data**

$$PX = Y$$

$$\text{cov}(Y) = \text{cov}(PX)$$

$$\text{cov}(PX) = \frac{1}{N-1}(PX)(PX)^T$$

$$\text{cov}(PX) = \frac{1}{N-1}PXX^TP^T$$

$$\text{cov}(PX) = P\left(\frac{1}{N-1}XX^T\right)P^T$$

$$\text{cov}(PX) = P\text{cov}(X)P^T$$

$$\text{cov}(PX) = P(V\Lambda V^T)P^T$$

$$P = V^T$$

$$\text{cov}(PX) = \Lambda$$

Some words about PCA

- PCA is “an orthogonal linear transformation that transfers the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (*first principal component*), the second greatest variance lies on the second coordinate (*second principal component*), and so on.”

Applications of PCA

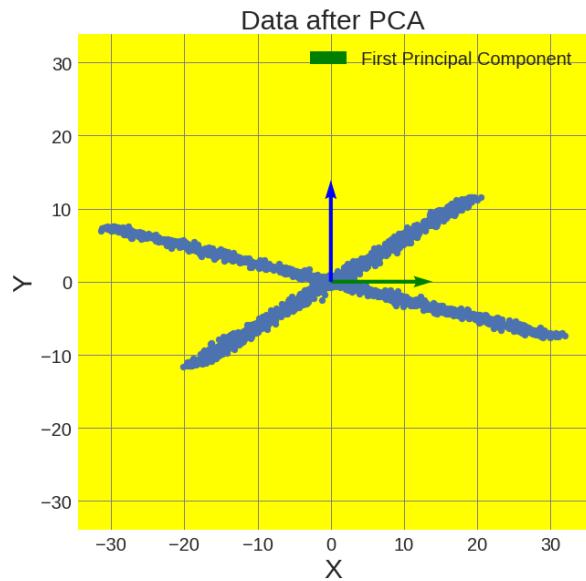
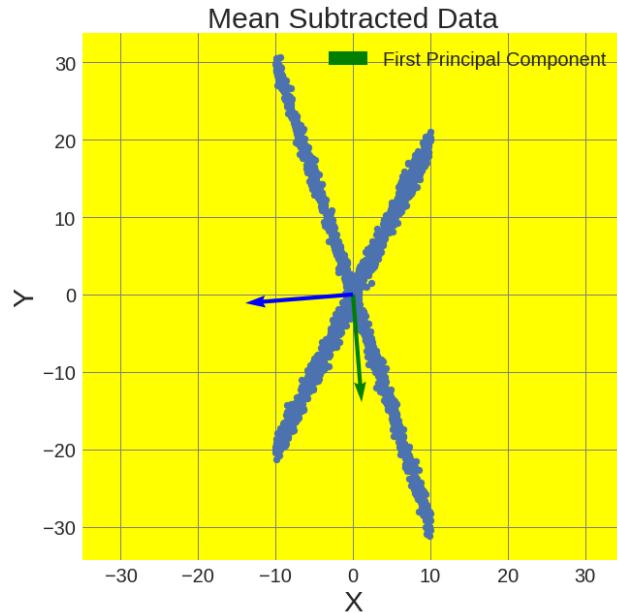
- Dimensionality Reduction
- Denoising
- Feature Extraction
- Image Compression
- EEG Analysis

Assumptions in PCA

- Linearity
- Large variance have important structure
- Principal components are orthogonal

When does PCA fail?

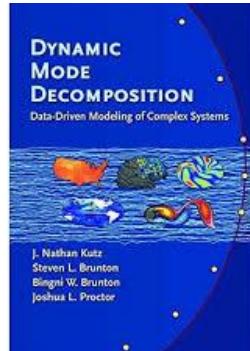
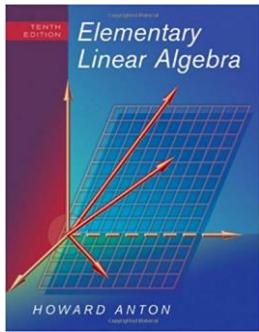
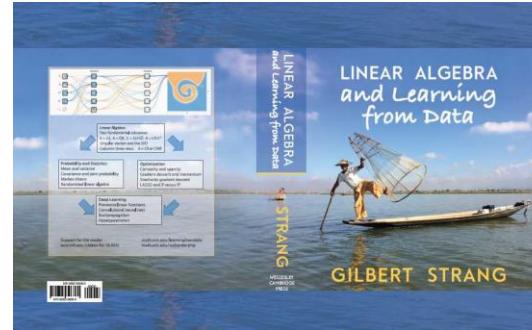
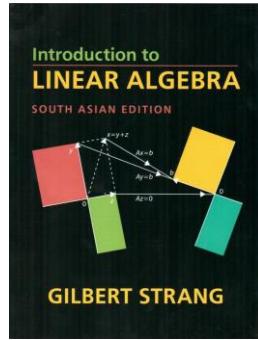
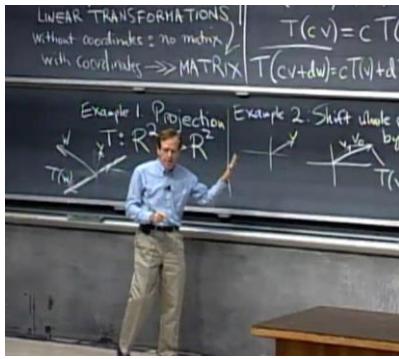
- Non-linearity
- Non-Gaussian
- Non-orthogonality



Ref: <https://arxiv.org/abs/1404.1100>

Interesting Materials

Prof. Gilbert Strang



Tutorial on PCA - [\(Click here\)](#)