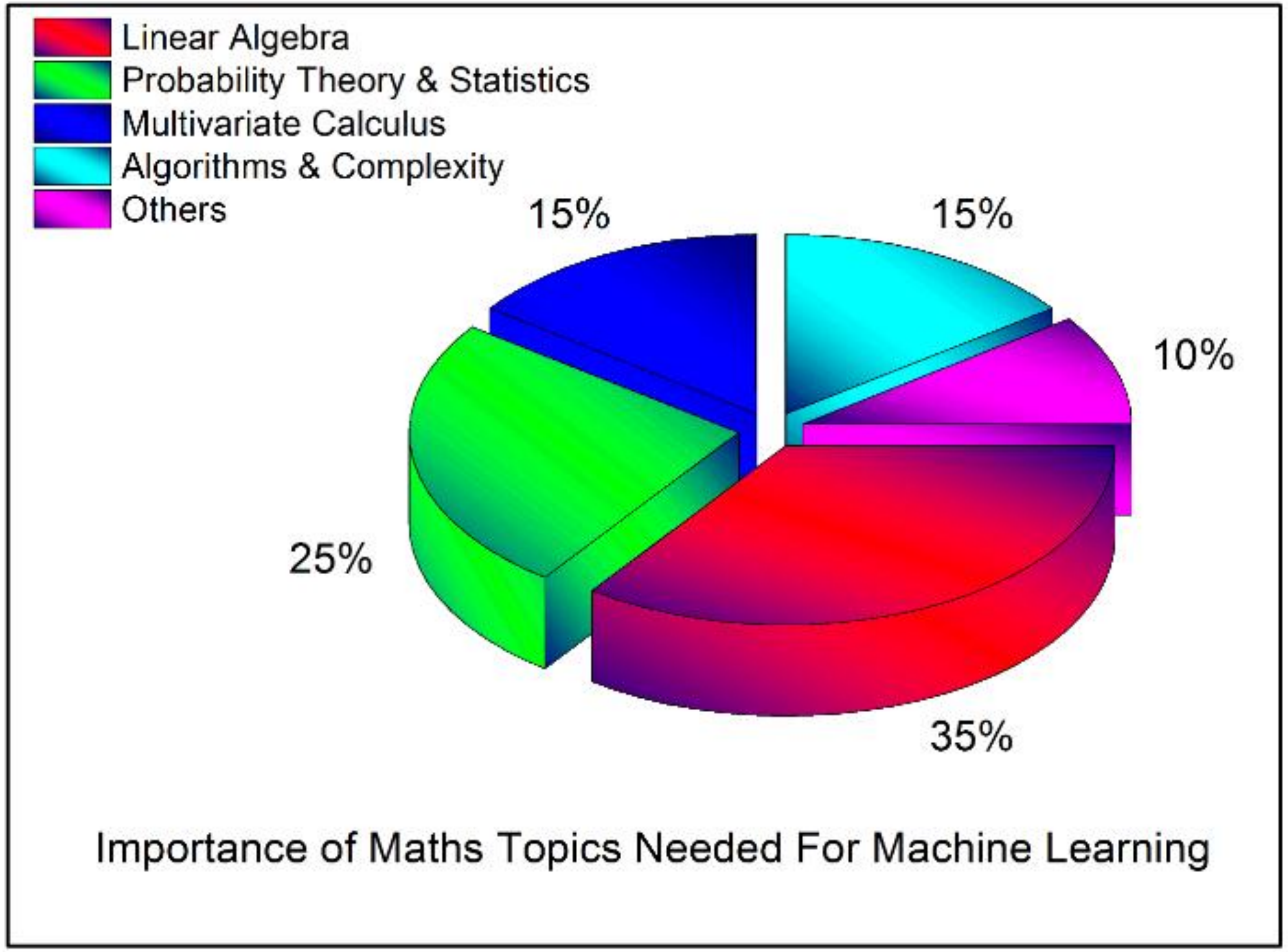
MATH
FOR
MACHINE LEARNING

Vidya Prasad

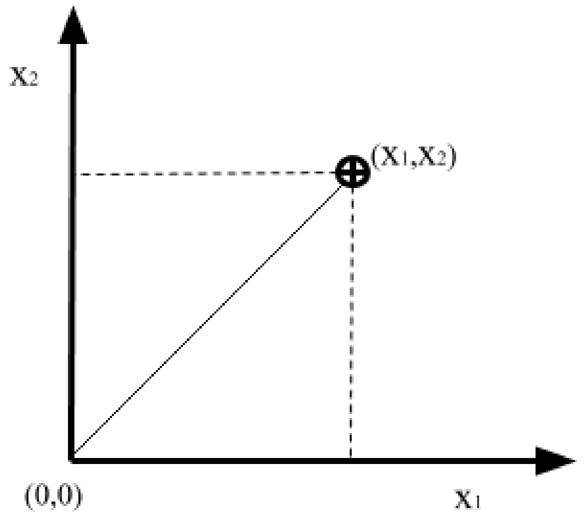

Linear Algebra
Probability Theory & Statistics
Multivariate Calculus
Algorithms & Complexity
Others
15%
15%
10%
25%
35%
Importance of Maths Topics Needed For Machine Learning

Linear Algebra

- Linear Algebra has Linear equations & functions which are represented using matrices and vectors.
- Algebra deals with scalars i.e. 1-Dimensional entities and
- Linear algebra deals with vectors and matrices (2 to N-Dimensional entities)

- **Why LA?**
- In machine learning , LA is used for the visualization of data
- Implementation of algorithmic solutions in computers is made easier.

Basic Terminologies



- **Vectors** : Vectors have both magnitude and direction.

For n-dimensional data,

$$\hat{x} = [x_1, x_2, \dots x_n]$$

$$x \in R^n, x_i \in R, 1 \leq i \leq 1$$

x_i: Element /Component

n : Number of elements in the vector or Dimension

Magnitude of x is found as:

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

This is also known as Norm-2 of the vector.

- **Vector Space:** A vector space V is **spanned** by a set of vectors say $\{a_1, a_2, \dots, a_n\} \in R^n$,then any vector in that space can be represented by the **linear combination** of these vectors i.e.

$$v = w_1 \begin{bmatrix} a_{11} \\ a_{12} \\ . \\ . \\ . \\ a_{1m} \end{bmatrix} + w_2 \begin{bmatrix} a_{21} \\ a_{22} \\ . \\ . \\ . \\ a_{2m} \end{bmatrix} + \cdots + w_n \begin{bmatrix} a_{n1} \\ a_{n2} \\ . \\ . \\ . \\ a_{nm} \end{bmatrix}$$

or $v = \sum_{i=1}^{n} w_i a_i = w_1 a_1 + w_2 a_2 + \cdots + w_n a_n, w_i \epsilon R$

- Since the set of vectors could span the vector space *V*, they are called the **basis** of *V*.

- **Linear Independence:**

A set of vectors is said to be linearly independent, if none of the vectors in the set can be represented as the linear combination of other vectors in the same set. Otherwise, the set of vectors is said to be linearly dependent.

- $a_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}, a_2 = \begin{bmatrix}2\\5\\7\end{bmatrix}$ and $a_3 = \begin{bmatrix}3\\7\\10\end{bmatrix}$ are linearly dependent on each other
- $b_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, b_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$ and $b_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$ are linearly independent of each other and span the space R^3.

- The number of basis vectors for a vector space V is the dimension of the vector space.

- **Inner Product/ Dot Product:**

Inner product of two vectors a,$b \in R^n$, denoted as $a^T b$ or $\langle a, b\rangle \in R$ is a scalar value or single number.

$$\langle a, b\rangle = [a_1, a_2, \dots, a_n]\begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{bmatrix} = \sum_{i=1}^{n} a_i b_i$$

T denotes transpose operation . Also $a^T b = b^T a$.

- **Orthogonality:**

Two vectors are said to be orthogonal when their dot-product is zero, i.e.

$$\langle a, b\rangle = 0$$

- **Matrices:**
- Collection of vectors or rectangular arrays.
- These vectors can be either arranged row or column wise.
- The size of the matrix is decided based on number of rows *m* and number of columns *n*. Consider a matrix A of dimension *m* x *n*

$$A = \begin{bmatrix} a_{11} & a_{12} \cdots & a_{1n} \\ a_{21} & a_{22} \cdots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{mn} \end{bmatrix}$$

- **Trace** of a Matrix is the sum of diagonal elements of a matrix

$$tr(A) = \sum_{I=1}^{N} A_{ii}$$

- **1-Norm** of a matrix is equal to the maximum of absolute column sum.
- **Infinity Norm** of a square matrix is the maximum absolute row sum
- **Euclidean Norm** or **Frobenius Norm** of a square matrix is the square root of the sum of all the squares of the elements.

  It treats the matrix like a vector

- Eg: For a Matrix $A = \begin{bmatrix} 2 & -1 & 3 \\ 4 & 5 & 2 \\ 2 & 3 & 1 \end{bmatrix}$,

- $\|A\|_1 = \max(8,9,6) = 9$
- $\|A\|_\infty = \max(6,11,6) = 11$
- $\|A\|_F = \sqrt{73}$

- **Rank of A Matrix**
- Rank (r) of a matrix basically refers to number of linearly independent row vectors or column vectors.
- Maximum number of linearly independent row vectors is same as maximum number of linearly independent column vectors.
- **Different types of matrices** are:
- Identity Matrix : Diagonal elements are 1 and non-diagonal elements are zeros.

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

- Zero Matrix: All elements are zeros.

$$Z = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- Diagonal Matrix: D= $\begin{bmatrix} 5 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & -3 \end{bmatrix}$

- **Scalar Matrix :** $S = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$
- **Upper Triangular Matrix:** $A = \begin{bmatrix} 5 & 9 & 3 \\ 0 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$
- **Lower Triangular Matrix: B**$= \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 4 & 9 & 5 \end{bmatrix}$
- **Invertible Matrices**
- If A is a square matrix of order *m*, and if there exists another square matrix B of the same order *m*, such that $AB = BA = I$, then B is called the inverse matrix of A and it is denoted by A^{-1}.
- A square matrix is invertible when A has a full rank.
- **Singular Matrices**
- A square matrix that does not have a matrix inverse.
- A matrix is singular if and only if its determinant is 0.
- Eg. $A = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix}$, det(A) = 3x2-6x1 = 0 $\Longrightarrow$A is a Singular Matrix

- For non-invertible matrices like rectangular matrices,
- If columns of matrix A are independent, we get left inverse of A
- If rows of matrix A are independent, we get right inverse of A
- This inverse is called Pseudo-Inverse, A^+.
- Consider a linear system of equations,

$$2x_1 + 3x_2 = 1$$
$$5x_1 + x_2 = 3$$

It can be rewritten as $\boldsymbol{Ax} = \boldsymbol{b}$ where $A = \begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 3 \end{bmatrix},\ A \epsilon R^{2x2}, x, b \epsilon R^2$

- The **Moore-Penrose pseudo-inverse(A^+)** is a generalization of the matrix inverse when the matrix may not be invertible. If A is invertible, then $x = A^{-1}b$ else, $x = A^+b$
  1. $AA^+A = A$
  2. $A^+AA^+ = A^+$
  3. $(AA^+)' = AA^+$
  4. $(A^+A)' = A^+A$
- If A is an mxn matrix where $m > n$ and A is of full rank ($= n$), then $A^+ = (A'A)^{-1}A'$

Eigen Values and Eigen Vectors

- For any nxn matrix A, scalars λ that satisfies A$x = \lambda x$,if $x \neq 0$ for any non zero vectors x are called as the eigenvalues λ and the corresponding non zero vectors x are called eigenvectors.
- Eg. For A = $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, taking A$x = \lambda x$,

  $(A - \lambda I)x = 0$ ⇒Characteristic Equation

  $|A - \lambda I| = 0 \Rightarrow (1 - \lambda)^2 - 4 = 0$

  $\lambda^2 - 2\lambda + 3 = 0 \Rightarrow$ Characteristic Polynomial

  to get $\lambda_1 = -1, \lambda_2 = 3$ which are the Eigen Values
- We substitute them in the Characteristic Equation to get the Eigen Vectors.
- Product of Eigen Values = Determinant of the matrix
- Sum of eigenvalues = Sum of diagonal elements = Trace(A)

Eigen Decomposition

- Matrix decomposition decompose a matrix into special type of matrices for easy manipulation in solving problems like linear equations.
- Defined only for square matrices
- Say A be a matrix with *n* Eigen values $\lambda_1, \lambda_2, \ldots, \lambda_n$ i.e. $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_n]^T$ and *n* Eigen vectors $\mathrm{V} = v_1, v_2, \ldots, v_n$, then A can be decomposed into

$$\boldsymbol{A = V\, diag(\lambda) V^T}$$

- In deep learning, we often deal with real symmetric metrics.
- **Real symmetric matrices** are **eigen decomposable** and the equation can be further simplify to:

$$\boldsymbol{A = Q\, \Lambda Q^T}$$

where, Q:orthogonal matrix composed of eigenvectors of A

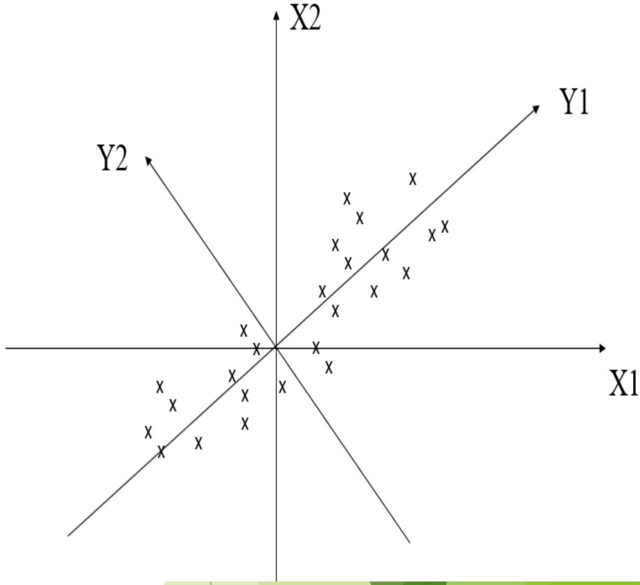
Λ:diagonal matrix

Principal Component Analysis

- It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences.
- Once the patterns are identified, the data is compressed by dimensionality reduction

1. Let $\bar{X}$ be the mean vector (taking the mean of all rows of a dataset)
2. Adjust the original data by the mean $X' = X - \bar{X}$
3. Compute the covariance matrix C of adjusted X
4. Find the eigenvectors and eigenvalues of C.
5. The Feature vector is formed using the Eigen vectors and this vector is used to define the new dataset by taking the transpose of the vector and multiplying it on the left of the original data set, transposed.

- Eigenvector with the highest eigenvalue is the principle component of the data set.



Singular Value Decomposition

- Any *m* x *n* matrix A can be represented as a product of three matrices:

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

- A is a $m \times n$ matrix. (Does not need to be a square matrix like eigen decomposition.)
- Left-singular vector: U is a $m \times m$ orthogonal matrix (the eigenvectors of AA^T)
- Right -Singular Vector: V is a $n \times n$ orthogonal matrix (the eigenvectors of $A^T A$)
- Singular values: D is a *m* x *n* diagonal matrix (square roots of the eigenvalues of AA^T and $A^T A$)
- SVD is a powerful but expensive matrix factorization method.
- In numerical linear algebra, many problems can be solved by representing A in this form

Positive definite or negative definite matrices

- If all eigenvalues of A are:
  - positive: the matrix is positive definite.
  - positive or zero: positive semi-definite.
  - negative: the matrix is negative definite.
- If a matrix is positive semi-definite, $x^T A x \geq 0$
- If a matrix is positive definite, $x^T A x = 0 \; i.e. \; x = 0$
- Positive definite or negative definite helps us to solve optimization problem.
- For positive definite matrices with non-zero x, the function is convex. i.e. it guarantees the existences of the global minimum.

- **Tensors:**
- A **tensor** is often thought of as a generalized matrix.
- The dimension of the tensor is called its *rank*.
- A tensor is a mathematical entity that lives in a structure and interacts with other mathematical entities.
- It is dynamic in nature
- Any rank-2 tensor can be represented as a matrix,

  but not every matrix is really a rank-2 tensor.