# MATHEMATICS FOR MACHINE LEARNING

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

Bachelor's degree in Mathematics


Jointly by,

| | |
|---|---|
| **ANN MARIA SAJAN** | **REG.NO:200021033165** |
| **ANEEGHA K A** | **REG. NO:200021033149** |
| **ANN MARY ANISH** | **REG. NO:200021033150** |
| **MARY BABU** | **REG. NO:200021033173** |
| **AKSHAY V A** | **REG. NO:200021033163** |
| **DEVIKA M R** | **REG. NO:200021033168** |

Under the guidance of

**MRS.ALSHA JOSEPH**

Department of Mathematics
Bharata Mata College
Thrikkakara, 2020-2023
Affiliated to Mahatma Gandhi University,
Kottayam

# ACKNOWLEDGEMENT

**We like to express our sincere thanks to Mrs. ALSHA JOSEPH, Associate Professor, Department of Mathematics , Bharata Mata College , Thrikkakara whose guidance helped us to complete this project work . We are also thankful to Dr. SEETHU VARGHESE, Head of The Department Of Mathematics for her helping attitude and valuable advices.**

**We are obligated to all our teachers of Department of Mathematics for their encouragement and co-operation for completing this work.**

**Place : Thrikkakara**

**Date :** 24-04-2023

***Ann Maria Sajan***
***Aneegha K A***
***Ann Mary Anish***
***Mary Babu***
***Akshay V A***
***Devika M R***

# CERTIFICATE

**This is to certify that the Dissertation entitled "MATHEMATICS FOR MACHINE LEARNING" submitted jointly by Miss. Ann Maria Sajan, Miss. Aneegha K A, Miss. Ann Mary Anish, Miss. Mary Babu, Mr. Akshay VA and Miss. Devika M R in partial fulfilment of the requirements for the Bsc. Degree in Mathematics is a bonafide record of the studies undertaken by them under my supervision of the Department of Mathematics, Bharata Mata College, Thrikkakara during 2020-2023. This dissertation has not been submitted for any other degree elsewhere .**

**Place : Thrikkakara**

**Date : 24- 04-2023**

**Mrs. ALSHA JOSEPH**
**Supervisor**

# DECLARATION

**We hereby declare that this project entitled "MATHEMATICS FOR MACHINE LEARNING" is a bonafide record of work done by us under the supervision of Mrs. ALSHA JOSEPH and the work has not previously formed the basis for any other qualification, fellowship , or other similar title of other university or board.**

**Place : Thrikkakara**

**Date : 24- 04 –2023**

***Ann Maria Sajan***
***Aneegha K A***
***Ann Mary Anish***
***Mary Babu***
***Akshay V A***
***Devika M R***

# TABLE OF CONTENTS

# CHAPTER 1: INTRODUCTION

- **Haar Bases and the associated Haar Wavelets, a vital tool in the processing of signals and computer graphics.**

- **Hadamard matrices, which are used in low rank estimation, processing of signals, and error fixing codes.**

- **Affine maps: These are typically dealt in a vague or ignored manner. Nevertheless, they are crucial to robotics and computer vision. The definition of affine maps is simple and elegant. The only requirement is to define affine combinations. Similar to how affine maps maintain affine combinations, linear maps preserve linear combinations.**

- **An affine function has a graph that is a line of equality and is made up of a linear function and a constant. y = Axe + c is the general equation for an affine function in one dimension. A linear transformation followed by a translation represents an affine transformation, which is demonstrated by an affine function.**

- **The subject of linear algebra is linear mixtures. To build new columns and palettes of numbers, arithmetic is used to arrays of numbers called matrices and columns of numbers called vectors. Linear algebra is an investigation of the planes, lines, spaces of vectors, and mappings required for linear transformations.**

- **An n-by-n matrix having 1 or -1 units and mutually orthogonal columns is known as a Hadamard matrix.**

- **An ML algorithm employs a variety of precise, probabilistic, and modern methodologies that enable computers to learn from the past and recognise challenging patterns in large, noisy, or complicated datasets.**

- **Artificial intelligence : The field of artificial intelligence integrates computer science with huge datasets to aid in problem-solving. It also encompasses the aspects of artificial intelligence known as machine learning and deep learning, which are frequently discussed in tandem.**

- **Deep Learning : A machine learning technique based on artificial neural networks in which input undergoes processing through many layers to extract features at progressively higher levels.**

- **Baye' s theorem : The chance of the second event given the first event multiplied by the probability of the first event is the conditional likelihood of an event dependent on the occurrence of another event.**

- **Confusion matrix : To explain how effectively a categorization system performs, a confusion matrix is utilised.**

- **To work with arrays, utilise the NumPy Python module. It additionally offers functions for working with Matrices, the Fourier Transform, and the field of Linear Algebra.**

# CHAPTER 2:
# MACHINE LEARNING

# Introduction of Machine Learning

**Since its introduction in the 1950s, machine learning has gained popularity thanks to developments in statistics, computer science, better datasets, and neural networks.**

- ## What is machine learning ?

**Machine Learning (ML) is a part of Artificial Intelligence (Al) that allow computers to 'self-learn' from training data and get better over time without having explicit programming. It is an automated procedure that allows robots to solve issues with little to no human involvement and acts in response to prior observations. Al is the more general idea of machines making decisions, picking up new abilities, and solving issues in a manner analogous to humans. While machine learning, a subset of artificial intelligence (Al), enables intelligent computers to learn new things on their own from data. Machine Learning can be used for processing enormous amounts of data and outperforms people by a wide margin.**

- ## What is the use of Machine Learning in Daily Life ?

**Enterprises need machine learning to comprehend consumer behaviour and create new goods.**

- **Applications of Machine Learning in Daily Life one by one**

**1) Traffic Aware (eg: Google Map):**

**Using two methods, it makes predictions about the state of the traffic, including whether it is clear, going slowly, or jam-packed.;**

- **The vehicle's real-time location as determined by sensors and the Google Maps app**
- **On similar days in the past, the average amount of time was needed.**

**2)Product Feedback :**

**Ads for the same product run across many channels.35% of Amazon's revenue comes from the recommendation of adverts based on search history, which is done via machine learning**

**3)Digital Personal Assistant :**

**Digital personal assistants can find helpful information with the aid of machine learning. Chatbots employ personal assistants to respond to inquiries and do information searches.**

**4) Autonomous Car :**

**Self-driving cars use machine learning, primarily NVIDIA's Unsupervised learning Algorithms. The € 750 billion in data from vehicles and drivers is crowdsourced via IOT sensors by NVIDIA's Deep Learning model.**

**5)Social Network :**
**Based on the facial detection and image recognition capabilities of DeepFace, machine learning is utilised to automatically tag people. Images posted to Facebook are given Alt Tags by Deep Face, which also recognises faces.**

**6)Transportation and Commuting :**
**Location is automatically determined by machine learning, which offers options based on previous experiences and patterns. Machine learning raises the accuracy of ETA predictions by 26%.**

**7)Google Translate :**
**Using Google Translate, tourists can connect with locals in several languages.**

**8) Image Recognition :**
**It is employed to identify individuals, locations, digital photos, etc. Automatic close friend tagging suggestions is a common application of picture recognition.**

**9) Email spam And Malware Filtering :**
**Every new email that we receive is automatically categorised as essential, normal, and spam. We consistently get critical emails in our inbox, important emails in our spam folder, and machine learning emails in our spam folder.**

**10) Online fraud Detection :**

**By identifying fraudulent transactions, Machine Learning tries to makes our online transactions safe and secure. Every time we make an internet purchase, there may be a number of ways for a fraudulent transaction to occur, among them include the use of fake accounts and identification cards, as well as the theft of money during a transaction. As a result, Feed Forward Neural Network assists us in identifying this by determining if the transaction is legitimate or fraudulent.**

## Mathematics for Machine Learning

- **Math is important for machine learning**

**In order to comprehend and use algorithms in a variety of applications, one must be an expert in mathematics. The machine learning process at every stage incorporates mathematical ideas, from picking the appropriate method to selecting the appropriate parameter. Other factors include complexity, bias in variance trade-off, and selecting an effective training time. Mathematical underpinnings are the basis of machine learning. To accomplish the Data Science project and resolve the Deep Learning use cases, mathematics is needed. The fundamental idea behind the algorithms is clarified by mathematics, which also shows why one algorithm is better than another. Even if you don't understand the reasoning behind how algorithms work, you can still create models.**

**Four machine learning pillars are used to tackle the majority of our real-world business issues. Additionally, many machine learning approaches leverage these pillars. They really are:**

- **Statistics**
- **Probability**
- **Linear Algebra**
- **Calculus**

## ❋ STATISTICS

**Everything revolves around Statistics. It is used to make inferences based on data. It focuses on statistical methods for compiling, displaying, analysing, and interpreting numerical data. Since statistics deal with enormous amounts of data and are essential to the expansion and development of an organisation, they are important in the field of machine learning. Censuses, Sampling, Primary and Secondary data sources, among other methods, can be used to collect data. We use this stage to help us define our goals so that we may go on to the following stages. The data collected is contaminated with errors, outliers, null values, noise, and other oddities. The data needs to be cleaned up and made into conclusions that can be put to use.**

**It is important to deliver the information in a suitable and condensed way. It is one of the most crucial procedures since it helps with understanding the insights and provides the framework for further data analysis. Examples of data analysis methods that make use of Central Tendency, Dispersion, Skewness, Kurtosis, Co-relation, Regression analysis and other methods. Making judgements from the collected data is part of the interpretation process because the statistics do not speak for themselves. Any truly effective machine learning starts with factual learning. One such factual tactic was regression, which was provided as an example. Statistics used in machine learning can be divided into two categories based on the type of analysis they do on the data. Inferential statistics and descriptive statistics.**

## ❋ PROBABILITY :

**Based on prior experiences, probability is the likelihood that a given occurrence will occur. It is used to predict the likelihood of future events in the field of machine learning.**
**The Bayes theorem explains how the conditional probabilities of events are connected. This theorem functions well with data samples that have some level of ambiguity and can be used to determine the 'Specificity' and 'Sensitivity' of data. The CONFUSION MATRIX is created using this theorem, which is essential.**
**A confusion matrix, which resembles a table, is used to assess how well machine learning models or algorithms perform. When building the ROC Curve from the supplied data, this is helpful for calculating the True Positive Rates, True negative rates, false negative rates, False positive rates, precision, Recall, F1-score, Accuracy, and Specificity.**
**The discrete and continuous varieties of probability distributions, as well as likelihood estimation functions, need to be given additional consideration.**
**A probabilistic machine learning algorithm called the Naive Bayes Algorithm makes the assumption that the input attributes are independent.**
**Numerous learning techniques are based on probability, including Nave Bayes and Bayesian Networks .**

## ❋ CALCULUS:

**This area of mathematics supports the investigation of quantity change rates. Its goal is to make machine learning algorithms and models perform better. Without understanding calculus, it is impossible to compute probabilities on data, and we are unable to predict realistic consequences from the data we gather. Calculus' foundational concepts include functions, integrals, limits, and derivatives. The two types of statistics are differential statistics and inferential statistics. Using back propagation techniques, it is utilised to train deep neural networks. To ascertain how the data evolves, differential calculus divides the input into discrete pieces. Inferential Calculus integrates (joins) the small bits to calculate how much there is. Calculus is mostly utilised to boost the accuracy of Machine Learning and Deep Learning Algorithms. It is used to offer efficient and timely solutions. Calculus is used in optimizers like Adam, Rms Drop, and Adadelta as well as calculus mostly for creating various Deep Learning and Machine Learning models. They support in the process of making better use of data by optimising it and creating better data outputs. Pattern learning can be implemented using calculus. The ML model uses a number of state and control combinations for analysis.**

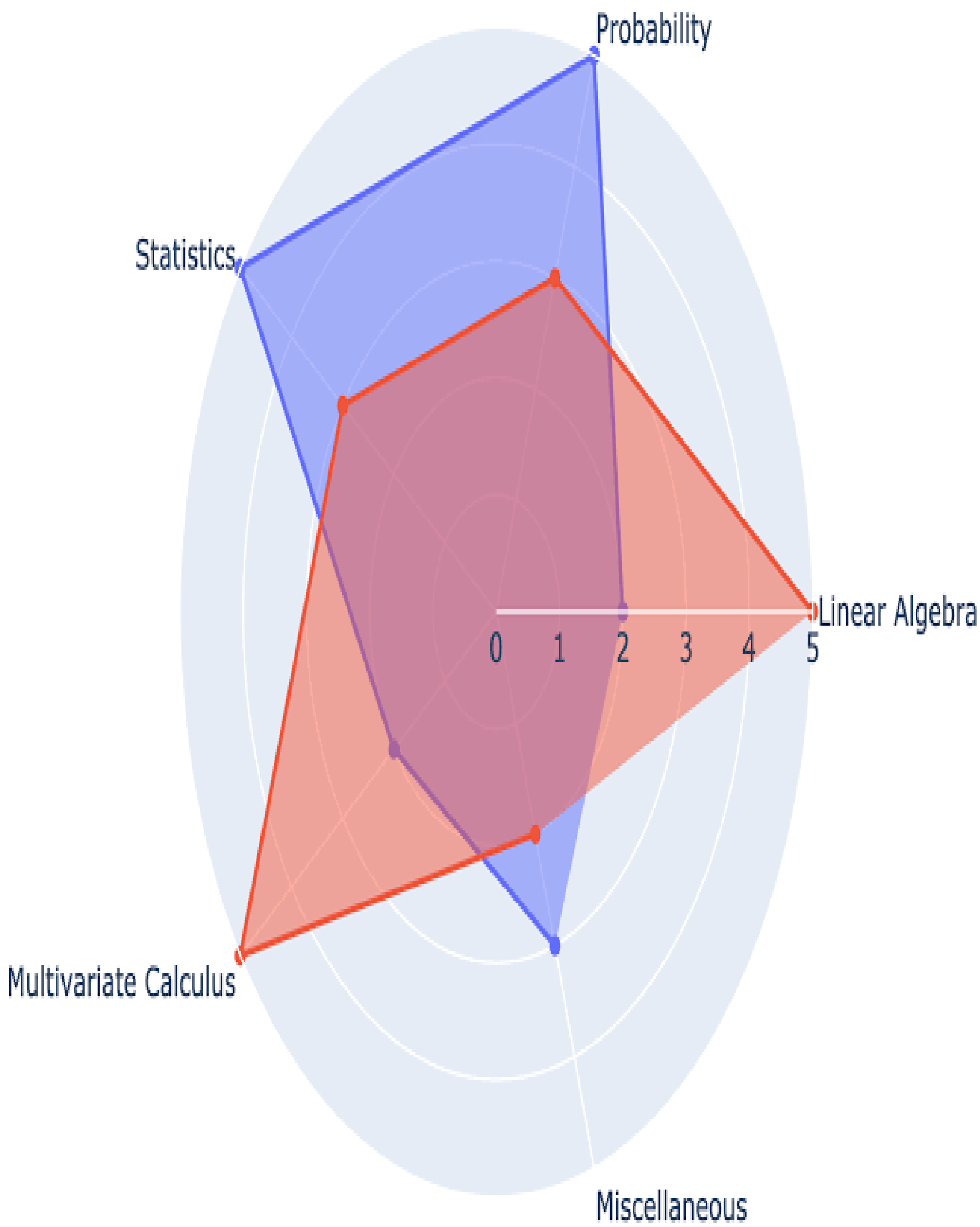
## ❋ LINEAR ALGEBRA :

**In linear algebra, computation is emphasised. Usage include Deep Learning and is crucial for understanding the underlying idea of Machine Learning. It gives us a clearer picture of how algorithms work in practise and helps us form more accurate conclusions. It mostly concentrates on matrices and vectors.**

- **A scalar is an integer with only one value.**
- **A vector is a numerical array with only one index (i.e., either Rows or Columns) and is stated as a row or column.**
- **A matrix is a two-dimensional array of numbers that may be accessed by both rows and columns as well as indices and keys.**
- **A tensor is a collection of integers with a variable number of axes and a grid arrangement in a specific sequence.**

**The Numpy module of the Python library is used to compute all of these numerical calculations on the data. The addition, subtraction, multiplication, and division operations that are performed on vectors and matrices by the Numpy library result in a meaningful value. The Numpy language is expressed as an N-d array. Without linear algebra, it is impossible to build machine learning models, maintain complex data structures, or perform matrix operations. The platform for presenting all the model results is linear algebra. Many Machine Learning methods, including Linear, Logistic Regression, SVM, and Decision Trees, are developed using linear algebra. Using linear algebra, we could even develop our own ML**

**algorithms. Data Scientists and Machine Learning Engineers use linear algebra to develop their own algorithms while working with data.**


Probability
Statistics
Linear Algebra
0
1
2
3
4
5
Multivariate Calculus
Miscellaneous

# CHAPTER 3:
# HADAMARD MATIRICS

# HADAMARD MATRICS

- Matrices Of Hadamard:

**Another well known family of matrices that are somewhat comparable to Haar matrices have entries +1 and -1.**

**DEFINITION: A Real n × n matrix H is a Hadamard matrix if $h_{ij} = \pm 1$ for all i, j in such a way that 1≤i, j ≤ n and if**

$$H^T H = nIn$$

**As a result the Hadamard matrix′s columns are pairwise orthogonal. The equation $H^T H = nIn$ demonstrates that H is invertible because it is a square matrix , thus we also get $H^T H = nIn$. Hadamard matrices include the following examples.**

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$H_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

**And**

$$H_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

**Find the positive numbers n for which a Hadamard matrix of size n exists is a logical question , yet curiously ,this one is still unsolved .The Hadamard conjecture states that there exists a Hadamard matrix of dimension n for each positive integer of the n=4k.**

**A necessary condition and many sufficient conditions are recognized.**

**Theorem;**
**if H is a $n \times n$ Hadamard matrix ,then n must either be 1,2 or 4k for any positive integer k .**

**Sylvester introduced a family of Hadamard matrices and proved that there are Hadamard matrices of dimensions $n=2^m$ for all $m\geq1$ using the following construction**

**( Sylvester,1867) The block matrix of measurements 2n is the case where H is a Hadamard matrix of dimension n,**

$$\begin{pmatrix} H & H \\ H & -H \end{pmatrix}$$

**is a Hadamard matrix.**

**If we start with**

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

**We obtain an infinite family of Sylvester-type symmetric Hadamard matrices. The Sylvester –Hadamard matrices $H_2$,$H_4$ and $H_8$ are represented by the letter $H2_m$. Hadamard provided instances of Hadamard matrices in 1893 for the numbers n=12 and n=20. Hadamard matrices are currently known for all n= 4k≤1000, with the expectation of n =668, 716 and 892.**

**Numerous applications of Hadamard matrices can be found in numerical linear algebra, signal processing, and error-correcting codes . For instance, there is a code based on H that can both detect and repair seven**

**faults in any 32-bit encoded block. In 1969 , a mariner spacecraft transmitted images back to the earth using this code.**

**The Walsh function are the piecewise affine functions plf(($H2_m$)I) associated with the $2_m$ rows of the Sylvester – Hadamard matrix H for any m≥0.**

**In order for the Walsh function Wal(k ,t) to be equivalent to the function plf(($H2_m$)I) associated with the row i of $H2_m$ that has k changes of signs between consecutive groups of +1 and consecutive groups of -1 , these 2 functions are often indexed by the integers 0,1,...,2m-1.For instance , the fifth row of $H_8$ , specifically**

(1 -1 -1 1 1 -1 -1 1)

**Has five blocks of +1 s and -1 s in a row ,four signs changes occur between these blocks, and as a result is a connected to Wal(4,t).**
**Walsh functions , in particular ,that correlate to the rows of H8 (from top to the bottom) are :**

**Wal(0,t) , Wal(7,t) , Wal(3,t) , Wal(4,t) , Wal(6,t) , Wal(2,t) , Wal(5,t)**

**Some publications refers to Sylvester-Hadamard matrices as Walsh Hadamard matrices due to the relationship between Walsh functions and these matrices. The $2^m$ Walsh functions are pairwise orthogonal for all m. The Countable set of Walsh Function Wal(k ,t) ,for all m≥0 and all k Such that , 0≤k≤2m-1 may be organised in such a way that it's an orthogonal Hilbert basis of the Hilbert space $L^2([0,1])$. Different techniques for**

**dimension reduction and low-rank matrix approximation use the Sylvester Hadamard matrix $H_2m$.**

**A specific type of structured dimension-reduction map is the subsampled randomized Hadamard transform.**
**An SRHT matrix is an l × n matrix of the type**

$$\phi=\sqrt{n/l}\ RHD \quad \text{where,}$$

**1)D is a random n × n diagonal matrix with independent random signs as its elements .**
**2) A normalised Sylvester – Hadamard matrix of dimension n is given by $H= n^{-1/2}H_n$**
**3) R is a uniformly distributed random l × n matrix that reduces an n-dimensional vector to coordinates.**

- **SUMMARY**

**The following is the list of the chapter's key ideas and findings :**

- **A brief overview of Haar wavelets and the Haar basis vectors.**
- **The [ tensor ] Kronecker product of matrices.**
- **Sylvester – Hadamard and Hadamard matrices.**
- **Walsh performs**

# CHAPTER 4:
# HAAR WAVELETS AND HAAR BASES

**We go through Haar matrices that are used in Computer Science and Engineering in this chapter:**

- **A fundamental tool in computer graphics and signal processing, Haar matrices and their accompanying Haar wavelets.**

## Introductory to Haar Wavelet Signal Compression:

**We start by taking a look at Haar wavelets in $R^4$. In audio and video signal processing, wavelets are particularly useful for condensing large signals into much smaller ones that nevertheless include sufficient information to render them visually and acoustically identical when played.**
**Consider the four Vectors $w_1$, $w_2$, $w_3$, and $w_4$ provided by**

$$w_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad w_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \quad w_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \quad w_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

**Because these vectors are pairwise orthogonal, their inner product is zero, proving that they are really linearly independent. Allow U = [$e_1$, $e_2$, $e_3$, $e_4$] be the canonical basis of $R^4$, and take W = [$w_1$, $w_2$, $w_3$, $w_4$] be the Haar basis.**
**Consider U = $R^4$'s canonical basis: $e_1$, $e_2$, $e_3$, and $e_4$. The formula for the basis matrix W = Pwu's transformation from U to W is**

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

**and we quickly discover that W's inverse is provided by**

$$W^{-1} = \begin{bmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

**Keep in mind that the first matrix in this product is $(W^TW)^{-1}$ and the second matrix in the previous product is $W^T$. This causes the vector v = (6, 4, 5, 1) over base U to become c = ($c_1$, $c_2$, $c_3$, $c_4$) W, over the Haar basis,**

$$\begin{bmatrix} c1 \\ c2 \\ c3 \\ c4 \end{bmatrix} = \begin{bmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$

**We first compute $c = W^{-1}v$ to convert a signal v = ($v_1$, $v_2$, $v_3$, $v_4$) to the coefficients of the signal c = ($c_1$, $c_2$, $c_3$, $c_4$) over the Haar basis.**

**Be aware that the signal's general median value, $c_1$, is equal to $(v_1+v_2+v_3+v_4)/4$.**

**The background of the image (or of the sound) is represented by the coefficient $c_1$.**

**Then, $c_2$ provides v's coarse details, $c_3$ provides v's first half's details, and $c_4$ provides v's second half's details.**

**The process of reconstructing the signal involves figuring out v =Wc. The secret to effective compression is to set some of the coefficients of c to zero, resulting in a reduced signal c, but retaining just enough essential information to have the reconstructed signal v = Wc look nearly as good as the original signal v.**

**Thus, the steps are:**

**input v → coefficients $c = W^{-1} v$→ compressed c^→ compressed v = Wc^.**

**Modern video conferencing is made possible by this type of compression strategy.**

**The reason for this has something to do with the reality that Haar wavelets are multiscale.**

**It turns out that there is a faster way to find $c = W^{-1}v$, without actually using $W^{-1}$.**

**Given the original signal v = (6, 4, 5, 1), we compute averages and half differences. We get the Coefficients as $c_3 = 1$ and $c_4 = 2$.**

**Then again we compute averages and half differences. The Coefficients as $c_1 = 4$ and $c_2 = 1$. Note that the original signal v can be reconstructed from the two signals and the signal on the left can be reconstructed from the two signals using first averages and first half differences. In particular, the data found using first averages and first half differences gives us :-**

$$5 + 1 = \frac{v1+v2}{2} + \frac{v1-v2}{2} = v_1$$

$$5 - 1 = \frac{v1+v2}{2} - \frac{v1-v2}{2} = v_2$$

$$3 + 2 = \frac{v3+v4}{2} + \frac{v3-v4}{2} = v_3$$

$$3 - 2 = \frac{v3+v4}{2} - \frac{v3-v4}{2} = v_4.$$

## The scaling characteristics of Haar Wavelets, Haar Bases and Haar Matrices

**The Method discussed above can be generalized to signals of any length $2^n$. The prior instance is equivalent to n = 2. Consider the scenario where n = 3. The matrix provides the Haar basis ($w_1$, $w_2$, $w_3$, $w_4$, $w_5$, $w_6$, $w_7$, and $w_8$).**

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

**This matrix's columns is orthogonal, making it clear that $W^{-1}$ = diag(1/8, 1/8, 1/4, 1/4, 1/2, 1/2, 1/2)$W^T$.**

**With the exception of the first, which is used for averaging, it appears that the following Haar basis vector, $w_2$, is the "Mother" of all the other basis vectors.**

**Indeed, in general, given**

**$w_2 = (1, \ldots, 1, -1, \ldots, -1)$,**

**the "scaling and shifting technique" is used to produce the remaining Haar basis vectors.**

**$w_2$, the scaling process generates the vectors**

**$w_3, w_5, w_9, \ldots, w_{2^j+1}, \ldots, w_{2^{n-1}+1}$,**

**such that $w_{2^{(j+1)}+1}$ is obtained from $w_{2^j+1}$ by forming two consecutive blocks of 1 and −1**

**of half the size of the blocks in $w_{2^j+1}$, and setting all other entries to zero. Observe that $w_{2^j+1}$ has 2^j blocks of 2^(n−j)**

**element. The Shifting Process be composed in shifting the blocks of**

**1 and −1 in $w_{2j+1}$ to the right by inserting a block of (k − 1)2^(n−j)**

**zeros from the left, with $0 \le j \le n-1$ and $1 \le k \le 2^j$ Be aware that we follow a tradition where j is used for scaling and k for shifting. Thus, we obtain the following formula for $w_2$**

**j+k:**

$$w^j_k(i) = \begin{cases} 0 & ,1 \le i \le (k-1)2^{(n-j)} \\ 1 & ,(k-1)2^{(n-j)} + 1 \le i \le (k-1)2^{(n-j)} + 2^{(n-j-1)} \\ -1 & ,(k-1)2^{(n-j)} + 2^{(n-j-1)} + 1 \le i \le k \times 2^{(n-j)} \\ 0 & ,k2^{(n-j)} + 1 \le i \le 2^n \end{cases}$$

**Followed by $0 \le j \le n-1$ and $1 \le k \le 2^j$. Of course $w_1 = (1, \ldots, 1)$.**

**If we slightly alter our indexing, allowing k to range from 0 to $2^{j-1}$, and utilising the index j instead of $2^j$, the aforementioned calculations appear a little better.**

## Transformation of a vector to its Haar Coefficients:

**Instead of utilising $W^{-1}$ to transform a vector u into a vector c of coefficients over the Haar basis using the matrix W to rebuild the vector u from its Haar coefficients c, we can utilise quicker techniques that use averaging and diffencing.**

**The series of vectors $u^0, u^1, \ldots, u^n$ is calculated as follows if c is a vector of Haar coefficients of dimension 2n:**

$$u^0 = c$$

$$u^{(j+1)} = u^j$$

$$u^{(j+1)}(2i-1) = u^j(i) + u^j(2^j + i)$$

$$u^{(j+1)}(2i) = u^j(i) - u^j(2^j + i) .$$

**for $j = 1,\ldots,n-1$ and $i = 0,\ldots, 2^j$. $u = u^n$ is the reconstructed vector (signal).**

**We compute the sequence of vectors $c^n, c^{(n-1)}, \ldots, c^0$ as follows**

$c^n = u$

$c^j = c^{(j+1)}$

$c^j(i) = [(c^{(j+1)}(2i-1) + c^{(j+1)}(2i)]/2$

$c^j(2^j + i) = [(c^{(j+1)}(2i-1) - c^{(j+1)}(2i)]/2$

**for $i = 1,\ldots,2^j$ and $j = n-1,\ldots,0$. $c = c^0$ is the vector over the Haar basis.**

**The conversion of a vector to its Haar coefficients for n = 3 is demonstrated here.**
**The Sequence u = (31, 29, 23, 17, - 6, - 8, - 2, - 4) is obtained:**

$C^3 = (31, 29, 23, 17, -6, -8, -2, -4)$

$$C^2 = \left(\frac{31+29}{2}, \frac{23+17}{2}, \frac{-6-8}{2}, \frac{-2-4}{2}, \frac{31-29}{2}, \frac{23-17}{2}, \frac{-6-(-8)}{2}, \frac{-2-(-4)}{2}\right) = (30, 20, -7, -3, 1, 3, 1, 1)$$

$$C^1 = \left(\frac{30+20}{2}, \frac{-7-3}{2}, \frac{30-20}{2}, \frac{-7-(-3)}{2}, 1, 3, 1, 1\right) = (25,-5,5,-2,1,3,1,1)$$

$$C^0 = \left(\frac{25-5}{2}, \frac{25-(-5)}{2}, 5, -2, 1, 3, 1, 1\right) = (10,15,5,-2,1,3,1,1)$$

**So c = (10,15,5,-2,1,3,1,1). Conversely, given c = (10,15,5,-2,1,3,1,1), we get the sequence**

$u^0 = (10,15,5,-2,1,3,1,1)$

$u^1 = (10+15, 10-15, 5, -2, 1, 3, 1, 1) = (25,-5,5,-2,1,3,1,1)$

**$u^2$ = (25+5,25-5,-5+(-2),-5-(-2),1,3,1,1) = (30,20,-7,-3,1,3,1,1)**

**$u^3$ = (30+1,30-1,20+3,20-3,-7+1,-7-1,-3+1,-3-1) = (31,29,23,17,-6,-8,-2,-4),**

**which gives back u = (31,29,23,17,-6,-8,-2,-4).**

# CHAPTER 5:
# CONCLUSION

**For machine learning aficionados and hopefuls, mathematics is a crucial subject to concentrate on, and a strong background in maths is required. Every concept you learn about machine learning has a direct or indirect mathematical connection, as does every tiny algorithm you create or use to solve a problem. The mathematic principles behind machine the foundational mathematics that we study in the eleventh and twelfth grades forms the basis for learning. At that moment, we gain theoretical knowledge, but in the realm of machine learning, we come across the real-world uses for the mathematics we previously studied . The best way to learn math concepts is to take a machine learning algorithm, find a use case, solve it, and understand the underlying arithmetic. Building various models for prediction, classification, audio or video recognition, etc., using machine learning is a hot trend in the field of computer science. Mathematical topics like linear algebra, probability, calculus, and statistics are required to build the machine learning model. We require a strong foundation in maths to develop machine learning solutions to real-world problems. The development of problem-solving skills is aided by a firm grasp of mathematical concepts.**

# TABLE OF SYMBOLS

| Symbol | Typical meaning |
|---|---|
| $a, b, c, \alpha, \beta, \gamma$ | Scalars are Lowercases |
| $x, y, z$ | Vectors are Bold Lowercases |
| A, C ,W, V | Matrices are Bold uppercase |
| $x^T$ , $A^T$ | Transpose of a Vector / Matrix |
| $A^{-1}$ | Inverse of a Matrix |
| $\langle x, y \rangle$ | Inner product of x & y |
| $x^T y$ | Dot Product of x & y |
| B = (b1, b2, b3) | (Ordered) tuple |
| B = [b1, b2, b3] | Matrix with horizontally nested column vectors |
| B = {b1, b2, b3} | Set of vectors (unordered) |
| $R^n$ | n-Dimensional Vector Space of Real numbers |
| ∅ | Empty set |
| D | No. of Dimensions; indexed by d = 1, . . . , D |
| N | No. of Data points; indexed by n = 1, . . . , N |
| $I_m$ | Identity Matrix of size $m \times m$ |
| $0_{m,n}$ | Matrix of Zeros of size $m \times n$ |
| $1_{m,n}$ | Matrix of Ones of size $m \times n$ |

| | |
|---|---|
| **dim** | **Dimensionality of Vector Space** |
| **det(A)** | **Determinant of A** |
| **rk(A)** | **Rank of matrix A** |
| **$\|\cdot\|$** | **Absolute value / Determinant** |
| **$\theta$** | **Parameter vector** |

## TABLE OF ABBERVIATIONS AND ACRONYMS

| ACRONYM | MEANING |
|---|---|
| **e.g.** | **for example** |
| **i.e.** | **this mean** |
| **i.i.d.** | **Independent, Identically distributed** |
| **PCA** | **Principal component analysis** |
| **SPD** | **Symmetric, positive define** |
| **SVM** | **Support Vector machine** |

# REFERENCES


* Marc Peter Deisenroth A. Aldo Faisal Cheng Soon Ong - MATHEMATICS FOR MACHINE LEARNING
* Alpaydin, Ethem. 2010. Introduction to Machine Learning
* Axler, Sheldon. 2015. Linear Algebra Done Right
* Adhikari, Ani, and DeNero, John. 2018. Computational and Inferential Thinking: The Foundations of Data Science