International Journal of Engineering, Science and Mathematics
Vol. 12 Issue 4, April 2023
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Linear Algebra for Machine Learning: Critical Concepts

Hansraj Gupta, Assistant Professor
Department of mathematics, Govt College Kaithal
Email: hansrajhctm@gmail.com

Abstract

Linear algebra is a branch of mathematics that deals with vectors, matrices, and linear transforms, and it plays an important role in Machine Learning algorithms, data analysis, and classification, among other areas. Machine Learning is used to give computers the ability to learn and make decisions on their own, a field that shares many similarities with mathematics, especially Linear Algebra, which is utilized in areas such as graphing, prediction, and regression. In this work, we examine the relationship between linear algebra and machine learning, and we provide a quick introduction to certain statistical ideas because this field is so inextricably linked to mathematics and statistics. Finally, we demonstrate how linear algebra, namely matrices and vectors, are used in popular machine learning methods like linear regression and support vector machine.

Keywords: Machine Learning, Linear Algebra, Application, Mathematics, Algorithms.

Introduction

The field of study known as machine learning focuses on making machines mimic human intelligence without being explicitly instructed to do so. Without even realizing it, we make use of machine learning often. Machine learning is used in a wide variety of applications, from identifying spam in emails to correcting spelling errors and even recommending videos to watch on YouTube, which is how you ended up here. Algorithms in machine learning are taught to complete certain tasks by being exposed to relevant data. This implies that as time goes on and new data becomes available, we can simply let our application discover patterns and adapt to the new information without having to rewrite it. Artificial intelligence, or "machine learning," is the branch of computer science that aims to give robots cognitive abilities similar to those of humans through the development of sensing, reasoning, acting, and adapting machines (Gennady Grabarnik, 2020). The concept of "deep learning" refers to a subfield of machine learning that takes its cues from how the human brain functions. A future where machines can learn and reason is being ushered in by machine learning. When dealing with a machine learning problem, one must go through the procedure of "Model Selection" to determine which model would work best. There are a number of variables to consider while choosing a model, including the dataset, the task, the type of model, etc. All of our efforts here will be directed at developing a novel mathematical machine model. Linear algebra is a branch of mathematics that deals with concepts like vectors, matrices, norms, and broadcasting (Anzt Hartwig, 2017).

Machine Learning

The analytical modeling process may be automated with the use of machine learning, which is a type of data analysis. Machine learning is a subfield of AI that seeks to automate as much of the learning, pattern recognition, and decision-making processes as possible. Tom Michel, a professor at Carnegie Mellon University, has offered one of several definitions of machine learning: Software that, depending on the T-task it's meant to perform and the P-function it's meant to fulfill, creates the E-experience. As a metric of P, performance increases when experience in E is gained. In this sense, we might claim that a machine has learnt if, for instance, it is able to increase its performance on a given job by drawing on its stored data. Similarly, data mining is the same as machine learning (Venkateswara Rao, 2009).

Machine learning systems receive multiple inputs and perform statistical analysis to generate outputs in a single interval, whereas in traditional computational methods, algorithms are written as a set of explicit commands to perform computation or problem solving for the purpose of running computers. Machine learning allows computers to automatically complete the Decision Making process on fresh incoming data by modeling it based on the sample data they were given.

Linear Algebra

One use of AI is machine learning, which gives computers the capacity to learn and develop on their own without being expressly programmed to do so. In order to create intelligent computer programs, we employ Machine Learning. To put it simply, linear algebra is an area of mathematics that has applications in the physical and social sciences and engineering. It's useful for learning about and using machine learning and algorithmic processes, especially those used for deep learning. An elementary machine learning algorithm built using existing machine learning expertise. Vectors and matrices will be the focus of linear algebra. All computations rely on particular kinds of vectors and matrices, thus it's crucial that you grasp them. The components of a feature vector are "features" of the item being studied (Vijayabarathi.S, 2014).

Linear Algebra Model

In order to define and assess the performance of a system, it is necessary to create a linear algebra model of the system, which involves identifying and describing the different cause and effect linkages between the system's parameters, decision variable, and state variables. As part of system modeling, identifying the system's parameters is essential. It is possible to estimate the system's performance criterion by intentionally generating events in the model using the decision variables and system parameters, and then tracking the resulting shifts in state variables.

Concept of Linear Algebra

- **Vectors:** As vectors are arrays of numbers in just one dimension, they are generally written with lowercase letters, italics, and boldface. An element in a vector may be retrieved using its index since the vectors are sorted. A vector represents a location in three-dimensional space. A two-dimensional matrix can be represented by a two-vector, a three-dimensional cube by a three-vector, and an n-dimensional tensor by an n-vector.
- **Matrix:** A matrix is a perfectly ordered set of data. Each piece of information in a matrix (which might be a number, symbol, or phrase) is referred to as an element. The numbers of columns and rows in a matrix might vary.
- **Norms:** Norms is a measure of the vector's simple (euclidean) distance from the origin, and it provides information on the vector's size. The standard notation for ||x||2 in machine learning is typically ||x||.
- **Vectorization:** It is notably useful for deep learning algorithms, which process large datasets in parallel, for which vectorization is a computational speedup strategy. One solution is to transform the data into matrices and vectors rather than doing scalar calculations in a for loop.

Machine Learning Process

- **Data collection:** With the issue description in mind, we now set out to gather relevant information from a variety of sources. Acquiring a sufficient amount of high-quality data is crucial.
- **Data preparation:** We take steps to clean up and clarify all of our information. in order for the ML model to quickly and correctly interpret the information. The data has to be divided into a training set and a test set.
- **Model selection:** Which model is best for our purposes will depend on the specifics of our data.
- **Model training:** Whatever data we have, we divide into a training set so that our ml model may learn from it and draw conclusions.
- **Prediction:** the process of estimating a missing or unknowable quantity from available information or historical trends.
- **Evaluation:** To determine how well our ML model is doing and whether or not it needs any adjustments, we put it through its paces using a test set of data.

Types of Machine Learning

Examples of each machine learning type below.

Supervised Learning: In supervised learning, an algorithm is trained using data that has been labeled by a human with the intended and expected outputs, as well as the inputs and outputs corresponding to them as pairs of input / Outputs. Predicting future occurrences from historical data is a popular use of supervised learning. For instance, this technique may be implemented in the previous few months or weeks to foretell changes in the stock market in the following weeks and months. This technique has also been used to identify spam sent by people who are not themselves spammers (Ali, 2011).

Unsupervised Learning: As unlabeled data is more widespread and ubiquitous than labeled data, unsupervised learning is of critical importance. In unsupervised learning, the input data lacks any tags, therefore the learning algorithm seeks for common characteristics and properties of the input data. Simply put, the purpose of unsupervised learning is to unearth previously unnoticed relationships between variables in unlabeled data. In reality, the system will be able to recognize unique data qualities and employ them to classify the data under consideration thanks to the application of unsupervised learning methods (Boyd, 2018).

Reinforcement Learning: By repeated iterations of trial and error, the algorithm in reinforcement learning gradually improves its predictions and subsequent judgments based on its experience. One type of RL algorithm is the Markov Decision Process. To illustrate this point, and as a useful example, let's imagine a robot footballer that learns how to take the best option to shoot in a given situation by situating itself in a variety of game-like scenarios, making judgments tailored to each, and then progressively rectifying its faults (Cunningham, 2015).

Research Methodology

Statistical analysis involves the systematic collection, organization, interpretation, and use of numerical data. While machine learning is a subset of the field of statistics, we will first review some basic statistical notions:

Correlation:

An integer between 0 and -1 is used to express the strength of the relationship between two variables; this coefficient is known as the correlation coefficient. If there is no relationship between two parameters, as indicated by a correlation coefficient of 0, then those parameters can be considered independent. And if the two parameters' correlation coefficient is more than zero, it indicates that shifts in the two parameters are intertwined. When the correlation coefficient is positive, it indicates that the two variables are positively correlated; when it is negative, the two variables are negatively correlated. This means that a rise in one causes a fall in the other and vice versa if the first one falls. In humans, for instance, there is a positive link between height and weight, so that as people get taller, they also gain weight.

Regression:

The word regression signifies "return" in its literal sense. The regression allows one to anticipate and describe changes in one variable based on changes in another when the two variables are closely connected.

Mean:

The median is a measure of central tendency in statistics that is calculated by dividing the total number of items in a collection by that number. The average is a set of statistics used to summarize information that has already been measured. Here is the formula used by mathematicians to determine the average:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

Median:

When the number of members of a statistical population is odd, the number in the middle of the two halves is determined by taking the mean of the means, represented by M, as a measurement of the degree of inclination toward the center. If there are an even number of people in the statistical population, though, the median is determined by averaging the middle two people. I feel the need to explain that when calculating the median, one must first arrange the data from least to greatest and then choose the middle value as the median.

Mode:

The most common information within a statistical sample is called the mode. As we've already mentioned, the mean and the median only give us a single figure to work with, but there might be several modes or perhaps none at all in a statistical community.

Variance and Standard Deviation:

Variance is a statistical measure of dispersion. What this means is that it is a numerical variance that demonstrates how the data is dispersed around the mean value through the following formula:

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$$

In other words, once we have the mean, we can compute the variance by taking the difference between each individual's value and the mean and raising it to the power of two.

Standard Deviation:

Similarly, the variance root may be used to calculate a statistic called standard deviation, which displays the average distance of data sets from the mean. Keep in mind that a small standard deviation implies that the data are relatively uniform and have little dispersion, whereas a big standard deviation shows that the data are significantly spread out. The standard deviation may be computed using the following formula:

$$SD = \sqrt{\frac{\sum (x - \bar{x}^2)}{n}}$$

It's important to remember that standard deviation gives useful information, since we can tell which members of the statistical community fall within the standard range, which fall over the standard limit, and which fall below the standard limit just by examining the SD number.

Results

Several different algorithms and methods are utilized in machine learning, with some of the most popular ones being mentioned here.

Linear regression algorithm:

As its mathematical foundation is so straightforward, linear regression has become one of the most used data modeling techniques. This type of regression may be used to make predictions about one variable given the value of another whenever a linear relationship can be established between them. The goal of establishing a linear connection between two variables is to demonstrate that a change in the value of the independent variable results in an equal and opposite change in the value of the dependent variable, as shown by the simple coefficient of the linear relationship.

If the value of one variable is superimposed on the value of another on a graph, and the resulting shape looks like a straight line, then the relationship between the two variables is linear. This diagram is also known as a distribution diagram. The following distribution graph only depicts the twenty individuals' measured heights and weights (Figure 1).

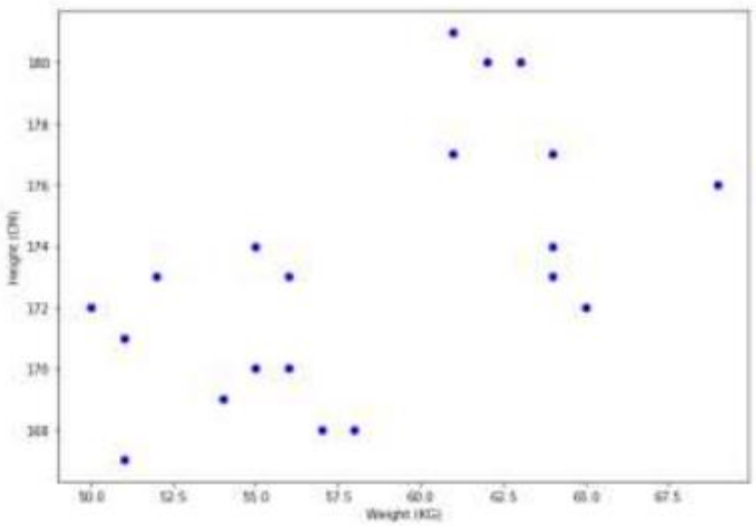

Figure 1: Statistics about the 20 participants' heights and weights

After establishing a linear connection between the two variables, we can apply a simple formula to extrapolate future readings of one variable from the other. Linear regression accomplishes just that—it determines the linear relationship between two variables by computing the coefficients of a linear relationship and then employing those coefficients to make predictions about the variables. Linear equations have the form $Y = w1 * X + w0$, where w1 is the line's slope and w0 is its width at the origin. While Y is the dependent variable and X is the independent variable in this connection, it is straightforward to derive X from Y and switch their roles as needed.

Mathematical calculations of linear equation coefficients:

We only need two points to determine a line's equation, but in practice, we often have thousands of data from which to select the optimal line. We need to determine the meaning of "best" before we can solve for the line equation's coefficients. We know there is a linear link between the square footage of a home and its monetary value, and thus we may utilize this information to our advantage. The data set may be viewed and downloaded at this location. To achieve this, we need to locate the optimal line depicting the cost distribution among the various housing components. Let's assume we do locate this line, and that the estimated price of every given property can be calculated by plugging in its area in this connection.

When we include in the available home sales data, also known as training data, we get an estimated selling price that differs from the real home price for each house. In linear regression, we try to find the line with the smallest discrepancy between the observed and predicted values. When we manually add up the differences between the real and estimated values, we get an irrational number since the positive and negative integers cancel each other out. Finding two coefficients W0 and W1 that are the same line coefficients can help us reduce this function. The Residual Sum of Squares is another name for this concept.

$$RSS(w_0, w_1) = \sum_{i=1}^{N} (y - [w_0 + w_1 x_i])^2$$

$$\sum_{i=1}^{N} a_i = a_1 + a_2 + \ldots + a_N$$

here,

$$a_i = (y - [w_0 + w_1 x_i])^2$$

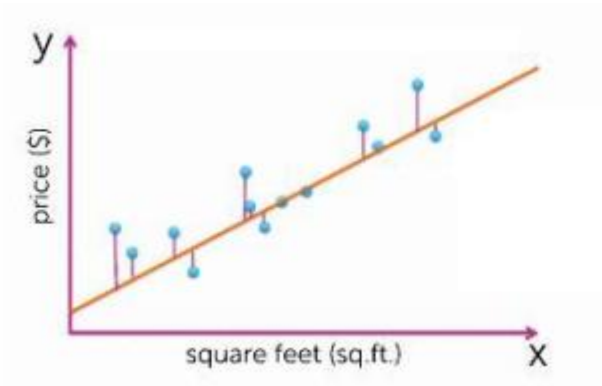

Figure 2: Residual sum of squares (RSS) diagram

The following graphic shows a three-dimensional representation of the quadratic function RSS(W0, W1), which yields varied results depending on the values of W0 and W1 for given x and y.

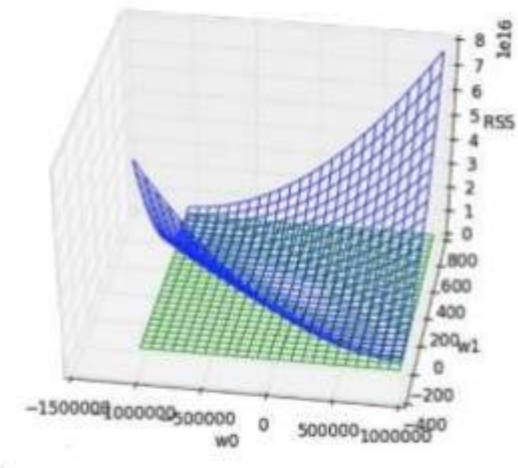

Figure 3: Minimum tangent plane in a three-dimensional visualization of the root-sum-squared

While the mathematical justifications for minimizing this function are outside the scope of this article, it is clear from the figure above that there is a global minimum and that this minimum occurs when the RSS changes to W0 and W1 are both zero. Hence, in practice, the gradient reduction approach is utilized, but in theory, the derivative of this function is obtained from both variables and set to zero. What follows from this is:

$$w_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$w_0 = \bar{y} - w_1\bar{x}$$

Here is the issue data, x, and the mean of our independent variable, y, where N is the total number of observations. Using the above formula, the correlation coefficient between x and y (denoted by W1) is the slope of the line, whereas the correlation coefficient between the mean of y and the product of x (denoted by W0) is the correlation coefficient between the two variables.

Interpretation of Linear Regression Coefficients:

While the slope and breadth of an interpretation source have clear meanings in a line's equation, that may not be the case in the data domain. If the area of the house being sold is 0, then the width of the origin will be equal to y. Consider a figure of $40,000. In general, we can assume that $40,000 is required to purchase a property, with the final price increasing with the square footage. The interpretation of the slope of our residential area is crystal obvious, and it is W1. The rate of change of y per unit increase in x is denoted by w1. If the slope of the line representing house sales is 280, for instance, this indicates that the price per square foot will increase to $ 280.

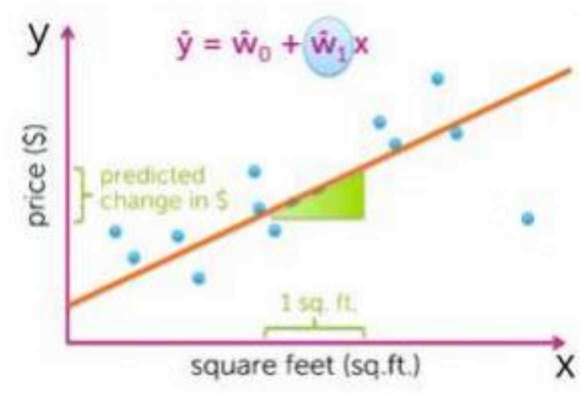

Figure 4: Diagram with little different slop

Multivariate regression:

By restricting the number of dependent variables to just one in the basic model for linear regression, multilinear regression may be thought of as an extension of linear regression that takes into account more than one independent variable.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

We assume that p is a random variable and that n is the number of observations of the dependent variable.

Support Vector Machine Algorithm

Within the realm of data categorization, SVM (support vector machine) is one of the most often used algorithms and techniques. Let's pretend we have a data collection where the gender split is 50-50. Users of an e-commerce platform might represent this data collection. With a portion of this data already labeled with people's genders, we can develop techniques to accurately predict the gender of the remaining people. We may enhance sales by targeting male and female customers differently based on their gender, which can be determined by analyzing customer demographic data. In data science, this step is known as classification. Assume that the gender, height, and hair length are three of the factors we need to figure out in order to provide a more complete picture of the issue at hand. Here is a bar chart showing the distribution of human height and length, broken down by gender using two squares (representing males) and a circle (representing females).

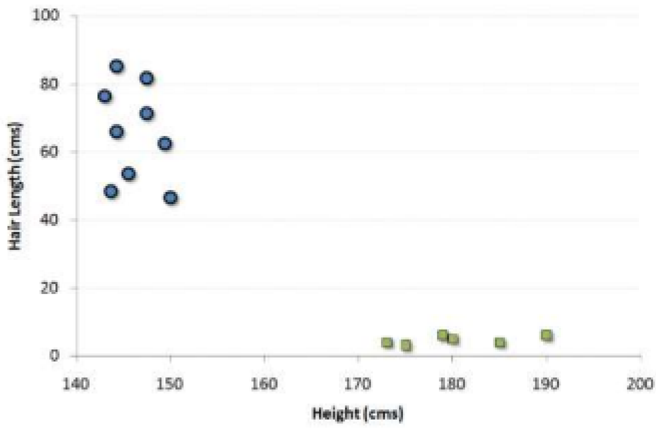

Figure 5: Height and hair length graph

With a height of 180 centimeters and a hair length of 4 centimeters, we would classify this individual as male.

Support vector machine implementation and nonlinear data distribution:

The aforementioned approach may provide the optimal machine for separating and classifying a data record if the data are linearly separable, but how to select SVM if the data are not linearly distributed (figure 6).

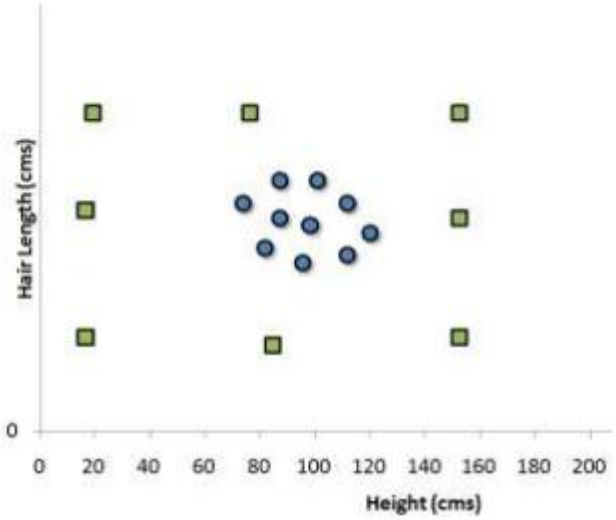

Figure 6: Linearly data distribution graph

As such, we need to use a kernel function to relocate the data to a new domain, one in which the information is readily divisible. The efficiency of the support vector machine is greatly improved by accurately determining this mapping function. Using the preceding illustration's conversion function, we can transform our data space into:

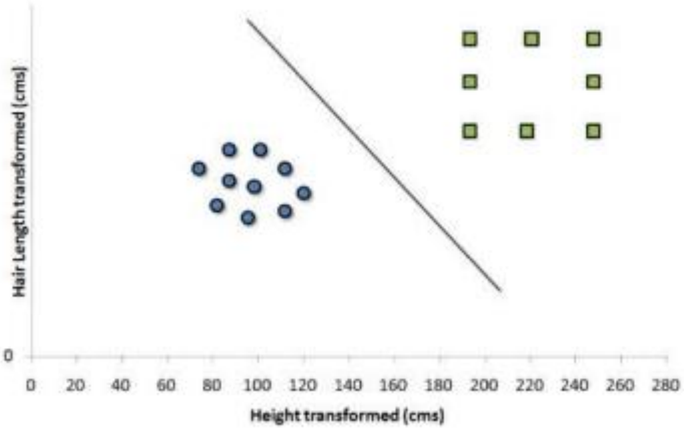

Figure 7: Mapping data using kernel function

As support vector machines are so computationally intensive, they also have a high memory footprint. However, this technique can't deal with discrete or otherwise non-numeric data, so you'll need to convert it. On the other hand, SVMs are based on a consistent theoretical framework, and their outputs are both universal and original. In recent years, support vector machines have emerged as the industry standard for making predictions in the data mining space.

Conclusion

the role that mathematics plays in the development and operation of Machine Learning Algorithms. This study describes the mathematical idea that inspired the development of Machine Learning algorithms, which were then utilized to successfully complete the set problem. We created metrics for data-sets and the vector is used for categorization and limits of datasets in n-dimensional space, along with Linear Algebra, Machine Learning, and Machine Learning techniques like linear regression and support vector machine. Experts in Machine Learning must require linear algebra. The field of statistics has used linear regression to characterize correlations between variables for quite some time. Predicting numerical values in less complex regression situations is a common use of this technique in machine learning. And When integrated with other machine learning techniques, such as the random forest approach, support vector machines become extremely potent algorithms for categorization and data separation. If the mapping functions are selected properly, this strategy is ideal for use in areas requiring extreme accuracy.

References

1. Ali, M & Pant, M 2011, ‘Improving the performance of differential evolution algorithm using Cauchy mutation’, Soft Computing, vol. 15, pp. 991–1007.
2. Alpaydin, Ethem. 2010. Introduction to Machine Learning. MIT Press.
3. Anzt Hartwig, Dongarra Jack, Flegar Goran and S. Enrique, "Quintana-Orti. Variable-Size Batched LU for Small Matrices and Its Integration into Block-Jacobi Preconditioning", 2017 46th International Conference on Parallel Processing (ICPP), pp. 91-100, 2017.
4. Arul, R., Ravi, G & Velusami, S 2013, ‘Chaotic self-adaptive differential harmony search algorithm based dynamic economic dispatch’, International Journal of Electrical Power & Energy Systems, vol. 50, pp. 85-96.
5. Bandurski, K & Kwedlo, W 2010, ‘A Lamarckian hybrid of differential evolution and conjugate gradients for neural network training’, Neural processing letters, vol. 32, no 1, pp. 31-44.
6. Barber, David. 2012. Bayesian Reasoning and Machine Learning. Cambridge University Press.
7. Baydin, Atılım G., Pearlmutter, Barak A., Radul, Alexey A., and Siskind, Jeffrey M. 2018. Automatic Differentiation in Machine Learning: A Survey. Journal of Machine Learning Research, 18, 1–43.
8. Boyd, Stephen, and Vandenberghe, Lieven. 2018. Introduction to Applied Linear Algebra. Cambridge University Press.
9. Cunningham, John P., and Ghahramani, Zoubin. 2015. Linear Dimensionality Reduction: Survey, Insights, and Generalizations. Journal of Machine Learning Research, 16, 2859–2900.
10. Gennady Grabarnik, Luiza Kim-Tyan, Serge Yaskolko, “Addressing Prerrequisited for STEM Classes Using an Example of Linear Algebra for a Course in Machine Learning”, The Twelfth International Conference on Mobile, Hybrid, and On-line Learning, 2020, pp. 21_26
11. Griewank, Andreas, and Walther, Andrea. 2008. Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation. SIAM.
12. SrinivasaRao.K On Study of dual ideals ,dual prime ideals of Pre A*-algebra in the proceedings of the Recent Trends in Dynamical systems and Mathematical Modelling,University of Madras , Chennai,25- 27,sep-2013,volume 18(2014),99-106.

13. Venkateswara Rao.J, Srinivasa Rao.K, Nageswara Rao.T., Srinivasa Rao R.V.N. Exploring Pre A*-Algebra as a New Paradigm", International Journal of Systemics, Cybernetics and Informatics, Jan 2009(14-19)
14. Vijayabarathi.S, SrinivasaRao.K , Boolean Algebras of Pre A*- algebra, International Journal of Mathematics Trends and Technology, Vol.7, Number 1, March 2014, pp 61-67