IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

5th National Conference on Science, Technology and Communication Skills – NCSTCS 2K22

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 9, Special Issue 2, September 2022

Linear Algebra: An Essential Kit for Machine Learning

Apurba Ghosh[1], Debangshu Bose[2], Soham Chakravorty[3] and Antika Mukherjee[4]

Assistant Professor, Basic Science and Humanities, Narula Institute of Technology, Kolkata, India[1,2]

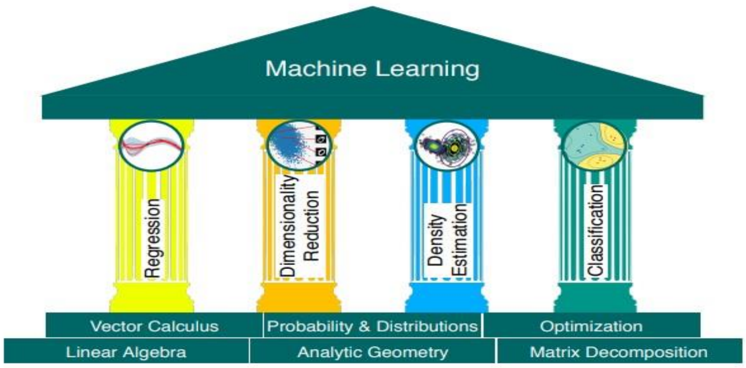
Student, Computer Science and Engineering, Narula Institute of Technology, Kolkata, India[3,4]

Abstract: Linear algebra is considered as a branch of Mathematics that involves the study of vector spaces, vectors, linear functions, and the system of linear equations, and matrices, linear functions in vector spaces. It's a field of study that has made a remarkable impact on other fields, such as statistics, as well as engineering and physics. In fact, it is eventually the mathematics of data. Machine learning is about designing algorithms that automatically extract valuable information from data. The emphasis here is on "automatic", i.e., machine learning is concerned about general-purpose methodologies that can be applied to many datasets, while constructing something that is meaningful. In this article our intention is to provide the mathematical background (especially Linear Algebra), applied to central machine learning problems that will help us to be precise about the concepts behind machine learning. We will focus on the need for mathematical conceptsby directly pointing out their usefulness in the context of fundamental machine learning problems.

Keywords: Linear Regression, Support Vector Machine, Vector, Independent Variable

I. INTRODUCTION

Machine learning is inherently data driven; data is at the core of machine learning. The goal of machine learning is to design general purpose methodologies to extract valuable patterns from data, ideally without much domain-specific expertise. While machine learning has seen many success stories, and software is readily available to design and train rich and flexible machine learning systems, we believe that the mathematical foundations of machine learning are important in order to understand fundamental principles upon which more complicated machine learning systems are built. Understanding these principles can facilitate creating new machine learning solutions, understanding and debugging existing approaches, and learning about the inherent assumptions and limitations of the methodologies we are working with.



Machine learning builds upon the language of mathematics to express concepts that seem intuitively obvious but that are surprisingly difficult to formalize. Once formalized properly, we can gain insights into the task we want to solve. One common complaint of students of mathematics around the globe is that the topics covered seem to have little relevance to practical problems. We believe that machine learning is an obvious and direct motivation for people to learn mathematics.

As applications of machine learning become widespread in society, we believe that everybody should have some understanding of its underlying principles.

The one of four pillars of machine learning (Fig. 1) we discuss in this article require a solid mathematical foundation. We represent numerical data as vectors and represent a table of such data as a matrix. The study of vectors and matrices is called linear algebra. The collection of vectors as a matrix is also described there. Given two vectors representing two objects in the real world, we want to make statements about their similarity. The idea is that vectors that are similar should be predicted to have similar outputs by our machine learning algorithm (our predictor). To formalize the idea of similarity between vectors, we need to introduce operations that take two vectors as input and return a numerical value representing their similarity.

Some operations on matrices are extremely decomposition useful in machine learning and they allow for an intuitive interpretation of the data and more efficient learning. We often consider data to be noisy observations of some true underlying signal. We hope that by applying machine learning we can identify the signal from the noise. This requires us to have a language for quantifying what "noise" means. We often would also like to have predictors that Draft (2022-01-11) of "Mathematics for Machine Learning".

II. BASIC ASPECTS OF LINEAR ALGEBRA

General aspects of linear algebra lie on a very simple concept: linearity. A function f is said to be linear if it maintains the equation $f(\alpha t_1 + \beta t_2) = \alpha f(t_1) + \beta f(t_2)$ where t_1 and $t_2$be any two points for the function. We often practice the term linear combination to describe any expression built from a set of variables by multiplying each variable by a constant and adding the results [1]. Most of the models used by scientists and engineers define linear relationships between quantities. Scientists, statisticians, engineers, business folk and politicians develop and use linear models to make sense of the systems they study. Actually, linear models are frequently used to model even more complicated (no-linear) phenomena. We can feel there are several decent reasons to use linear models and one of that is linear models are very respectable at approximating the real world. Linear models for nonlinear phenomena are termed as linear approximations.

In order to describe nonlinear phenomena linear models can also be united with nonlinear transformations of the model's inputs or outputs. These techniques are frequentlyapplied in machine learning: kernel methods are arbitrary non-linear transformations of the inputs of a linear model, and the sigmoid activation curve is used to transform a smoothlyvarying output of a linear model into a hard yes or no decision [2].

Equation of the tangent to the curve $f(x)$ at x_0 is described by $T(x) = f(x_0) + (x - x_0)f'(x_0)$.Here $f'(x_0)$ is the slope of this line passing through the point $(x_0, f(x_0))$ and the tangent line $T(x)$can be seen as approximation of the function $f(x)$near $x = x_0$. Modelling of nonlinear phenomena can be described as a multivariable generalization of this idea while using linear algebra.

One of the main reasonsthat linear models are widely used is linear models are informal to describe mathematically, and easy to "fit" to real-world systems. One can obtain the parameters of a linear model for a real-world system by analysing its behaviour for relatively few inputs. For example, at an art event, one person enters a room organised with multimedia setup. There is a drawing canvas on a tablet computer projected on a big screen andanything drawn on the tablet will instantly appear on the bigscreen. More interestingly, the user interface on the tablet screen doesn't give any indication about how to hold the tablet "right side up." Then what would be the fastest way to find the accurate orientation of the tablet so the drawing will not appear upside-downor rotated? This kind of situation is directly equivalent to the jobs scientists encounter every day when trying to model real-world systems. The canvas on the tablet and the wall projection can be correspond to two-dimensional input and output spaces respectively. We're eager to finda linear transformation T that relates the pixels of the tablet screen (the input space) to colored dots on the wall (the output space). It is evident that if the transformation T is a linear transformation, one can learn its parameters very rapidly.

If we describe every pixel in the input space with a point (x, y) on two-dimensional xy-plane and every point on the wall with another point $(\bar{x}, \bar{y})$in another co-ordinate system, then an unknown transformation will map pixel co-ordinates to wall co-ordinates: $(x, y) \to (\bar{x}, \bar{y})$.

To reveal how the linear transformation transforms (x, y)coordinates to $(\bar{x}, \bar{y})$coordinates, we will follow a procedure as follows. Initially we place a dot in the lower left corner of the tablet to signify the fixed point (0,0) of the xy −coordinate system.Perceive the place where the dot appears on the wall—we'll denote this location as the origin of anothercoordinate system. Further, take a short horizontal swipe on the screen to denote the x direction (1,0) and observe the transformed point$T(1,0)$that appears on the wall. Finally, make a vertical swipe in the y-direction (0,1) and see the transformed $T(0,1)$ that appears on the wall. By observing how the xy coordinate system is correspondent to the $\bar{x}\bar{y}$coordinate system, we can configure which orientation one must need for tablet to hold for appearing the drawing upright projected on the wall. For complete characterization of the linear transformation T it is enough to know the outputs of T for all directions in its input spaces [2].

III. FOUNDATION OF MACHINE LEARNING

Machine learning (ML) is a branch of artificial intelligence (AI) that helps computers to "self-learn" from training data and improve over time, without being explicitly programmed. Machine learning algorithms are able to detect patterns in data and learn from them, in order to make their own predictions.While artificial intelligence and machine learning goes hand in hand, they are two different concepts. AI is the broader concept – machines making decisions, learning new skills, and solving problems in a similar way to humans – whereas machine learning is a subgroup of AI that enables intelligent systems to autonomously learn new things

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

5th National Conference on Science, Technology and Communication Skills – NCSTCS 2K22

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 9, Special Issue 2, September 2022

from data.Machine learning can also be reallyuseful to handle massive amounts of data and can perform much more accurately than humans [3]. And thus, we can also save time and money on tasks and analyses, like solving customer pain points to improve customer satisfaction and etc.

There are mainly five types of machine learning: -

1. Supervised learning: Supervised learning algorithms make forecast based on labeled training data. Each sample comprises of an input and a desired output. A supervised learning algorithm analyzes this sample data and makes an inference – basically, an educated guess when determining the labels or unseen data.
2. Unsupervised learning: Unsupervised learning algorithms uncover insights and relationships in unlabeled data. In this case, models are given input data but the desired outcomes are not known, so they have to make intervention based on circumstantial evidence, without any training. The models are not trained with the “correct answer,” so they must find patterns on their own.
3. Semi-supervised learning: In semi-supervised learning, training data is split into two. A small amount of labeled data and a larger set of unlabeled data.In this case, the model uses labeled data as an input to make inferences about the unlabeled data, providing more accurate results than regular supervised-learning models.
4. Reinforcement learning (RL): It is concerned with how a software agent ought to act in a situation to maximize the reward. In short, reinforced machine learning models attempt to determine the best possible path they should take in a given situation. They do this through trial and error. Since thereis no training data, machines learn from their own mistakes.
5. Deep learning: Deep learning is constructed from Artificial Neural Networks (ANN), a type of computer system that copy the way the human brain works. Deep learning algorithms are built with multiple layers of interconnected neurons, allowing multiple systems to work together simultaneously [4-6].

Basically, the machine learning process involves three steps:

1. Feed a machine learning model training input data. In our scenario, it could be customer comments from social media or customer service data.
2. Desired output is tagged with the training data. In this case, we have to tell the sentiment analysis model whether each comment or piece of data is Positive, Neutral, or Negative. The model modifiesthe training data into text vectors – numbers that represent data features [7].
3. Test and run your model by feeding it testing (or unseen) data. Algorithms are trained to associate feature vectors with tags based on manually tagged samples, then predictions were made while processing the unseen data.

Machine learning is already widely used in finance, healthcare, hospitality, government, and beyond. Businesses are beginning to look at the profit of using machine learning tools to improve their processes, gain valuable insights from unstructured data, and automate tasks that would otherwise require hours of boring manual work [8-9].

For example, machine learning is used in apps like Zomato to estimate optimum times for drivers to pick up food orders, while Spotify by using machine learning offers personalized content and personalized marketing. And Dell uses machine learning text analysis to save hundreds of hours analyzing large number of employee surveys to listen to the voice of employee (VoE) and improve employee satisfaction.

How do you think Google Maps predictions in traffic and Netflix creating personalized movie recommendations, even informs the creation of new content? By using machine learning.

IV. CENTRAL MACHINE LEARNING PROBLEMS

Regression is a fundamental problem in machine learning, and regression problems appear in a diverse range of research areas and applications, including time-series analysis (e.g., system identification), control and robotics (e.g., reinforcement learning, forward/inverse model learning), optimization (e.g., line searches, global optimization), and deep learning applications (e.g., computer games, speech-to-text translation, image recognition, automatic video annotation). It is also a key ingredient of classification algorithms.

Linear regression is used to predict the value of a dependent variable (y) based on a given independent variable (x). As a result, this regression technique determines a linear relationship between x (input) and y. (output). As a result, the name Linear Regression was chosen. In the diagram above, X (input) represents work experience and Y (output) represents a person's salary. Our model's best fit line is the regression line.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

5th National Conference on Science, Technology and Communication Skills – NCSTCS 2K22

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 9, Special Issue 2, September 2022

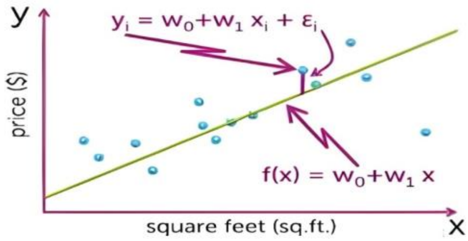


Fig. 2 Relationship between area of the house and their prices

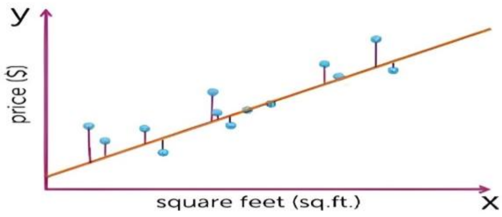


Fig 3. Sum of Squares Diagram

Now, if we are given with the data of selling homes and for each house, an estimated selling price that can be obtained that will be unlike from the actual home price. Our purpose in linear regression is to determine a line which has the smallest difference between the available data and the estimated data.

So, if it is possible to construct a function that displays the variance between the actual values and the estimated values, like the sum of the squares of these differences, then we can initiate a step to find two coefficients w_0 and w_1 that are the same line coefficients to minimalize this function, denoted by SOS and defined as,

$SOS(w_0, w_1) = \sum_{j=1}^{M}(y - (w_0 + w_1 x_j))^2 = \sum_{j=1}^{M} b_j = b_1 + b_2 + \cdots + b_M$, where

$b_j = (y - (w_0 + w_1 x_j))^2$, and M is the number of available data.

$SOS(w_0, w_1)$ is a quadratic function and for different values of w_0 and w_1, having x and y, it produces different outputs that we can observe in the following figure in three-dimensional form:

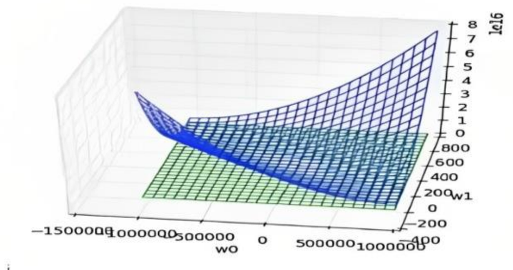


Fig. 4 3D plot of SOS with tangent plane at minimum

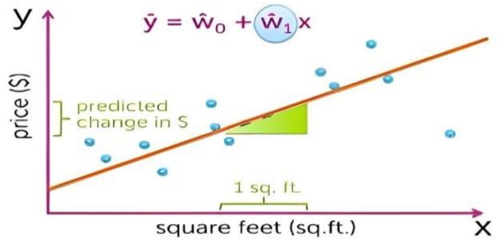


Fig 5. Diagram with little different slop

We minimize the function $\sum_{j=1}^{M}(y - (w_0 + w_1 x_j))^2$over all possible w_0 and w_1.

Mathematical arguments to minimize this function are beyond the reach of this discussion but if we look at the figure above, we observe that there has a global minimum and furthermore this minimum occurs when the functional value changes to w_0 and w_1 are zero (To observe this, suppose we are moving down the blue curve, which is the minimum function. At the bottom, with slight variations in each of the two variables, our displacement will be negligible compared to the rest). Therefore, the derivative of the above function is taken from both variables and set to zero. The result is:

$w_1 = \frac{\sum(x_j - \bar{x})(y_j - \bar{y})}{\sum(x_j - \bar{x})^2}$, and $w_0 = \bar{y} - w_1 \bar{x}$

In the equation of a line, the slope and width of the source of interpretation have a certain interpretation, but in the data domain, the width of the source may not be interpretable. Looking at the example of a home sale, if the area of the house is zero, y will be equal to the width of the origin. Suppose that the number is \$ 40,000. We can say that at least \$ 40,000 is needed to buy a home, and then the price will increase depending on the size of the home. Our house area has a clear slope of interpretation or $e_1$interpretation. e_1 denotes the rate of change in y for a unit x increase. In the example of home sales, for example, if the slope of the line is 280, it means that the price will rise to \$ 280 per square foot.

Linear regression can also be used to provide better insights by revealing patterns and relationships that your business colleagues may have previously seen and thought they understood. For example, analysing sales and purchase data can help you identify specific purchasing patterns on specific days or at specific times. Regression analysis insights can help business leaders predict when their company's products will be in high demand.

V. SUPPORT VECTOR MACHINES

Support Vector Machine, or SVM, is a popular Supervised Learning algorithm that is used for both classification and regression problems. However, it is primarily used in Machine Learning for Classification problems.

The SVM algorithm's goal is to find the best line or decision boundary for categorising n-dimensional space so that we can easily place new data points in the correct category in the future. A hyperplane is the best decision boundary.

SVM selects the extreme points/vectors that aid in the creation of the hyperplane. These extreme cases are referred to as support vectors, and the algorithm is known as the Support Vector Machine [10]. Consider the diagram below, which shows two distinct categories separated by a decision boundary or hyperplane:

Example: SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat.

SVM can be of two types:

- Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

The advantages of support vector machines are:

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

The disadvantages of support vector machines include:

- If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.
- SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation

As we have seen, SVMs depends on supervised learning algorithms. The aim of using SVM is to correctly classify unseen data. SVMs have a number of applications in several fields.

Some common applications of SVM are-

- Face detection – SVMs classify parts of the image as a face and non-face and create a square boundary around the face.
- Text and hypertext categorization – SVMs allow Text and hypertext categorization for both inductive and transudative models. They use training data to classify documents into different categories. It categorizes on the basis of the score generated and then compares with the threshold value.
- Classification of images – Use of SVMs provides better search accuracy for image classification. It provides better accuracy in comparison to the traditional query-based searching techniques.
- Bioinformatics – It includes protein classification and cancer classification. We use SVM for identifying the classification of genes, patients on the basis of genes and other biological problems
- Protein fold and remote homology detection – Apply SVM algorithms for protein remote homology detection.
- Handwriting recognition – We use SVMs to recognize handwritten characters used widely.
- Generalized predictive control(GPC) – Use SVM based GPC to control chaotic dynamics with useful parameters.

Thus, we conclude that the SVMs can not only make the reliable prediction but also can reduce redundant information. The SVMalso obtained results comparable with those obtained by other approaches.

VI. CONCLUSION

In this paper we discussed Linear Algebra, Machine Learning and Machine Learning algorithms like linear regression and support vector machine which explicitly have linear relationships to independent and non-independent variables. Linear algebra is that kind of branch in Mathematics Machine Learning experts can't live without. One will never evolve from amateur to good specialist without mastering this field of Mathematics. Basically, Mathematics is not the key to success, but in severalcases, it helps, and in a fewlike developing some deep learning models, it is essential. Linear algebra ideas and methods are easiest to comprehend, it will help to understand abstract concepts the best. Once we can see that the deeper you dive into machine learning, the more linear algebra we will see there. Linear regression is an old method from statistics for describing the relationships between variables. It is regularly used in machine learning for predicting numerical values in simpler regression problems. And Support vector machine are very powerful algorithms in classification and data separation, especially when combined with other machine learning methods such as random forest method. This method works great for places where we need very high precision data, provided that we choose the mapping functions correctly, it works very well.

REFERENCES

[1] O. Kharkovyna, "Mathematics for AI: Linear Algebra and How to Understand It Better," 12 10 2019. [Online]. Available: https://towardsdatascience.com/mathematics-for-ailinear-algebra-and-how-to-understand-it-better63b430999069.
[2] I. Savov, No bullshit guide to linear algebra, Montréal, Québec, Canada: Minireference Co, 2016.
[3] https://dev.to/duomly/the-best-programming-language-for-artificial-intelligence-and-machine-learning-e7b
[4] Sas, "machine learning," 7 11 2019. [Online]. Available: https://www.sas.com/en_us/insights/analytics/machin e-learning.html.
[5] D. Dwivedi, "Machine Learning For Beginners," towardsdatascience, 7 5 2018. [Online]. Available: https://towardsdatascience.com/machine-learning-forbeginners-d247a9420dab.
[6] J. Brownlee, "Supervised and Unsupervised Machine Learning Algorithms," https://machinelearningmastery.com/supervised-andunsupervised-machine-learning-algorithms/
[7] S. Mishra, "Unsupervised Learning and Data Clustering," 20 5 2017, https://towardsdatascience.com/unsupervisedlearning-and-data-clustering-eeecb78b422
[8] Statisticssolutions, "Correlation (Pearson, Kendall, Spearman)," https://www.statisticssolutions.com/correlationpearson-kendall-spearman/.
[9] M. D. Okoi, "Popular Programming Languages on GitHub," fossmint, 30 4 2019. [Online]. Available: https://www.fossmint.com/popular-programminglanguages-on-github/.
[10] N. Heath, "top 10 programming languages for machine learning," techrepublic, 2. Available: https://www.techrepublic.com/article/github-the-top10-programming-languages-for-machine-learning/.