Application of Linear Algebra in Machine Learning

Sahar Halim

Faculty of Computer Science, Balkh University, Balkh, Afghanistan

---

Abstract - *Linear algebra is a part of mathematic which include vectors, matrixes and linear transform that has applicable on different fields such as physic, engineering and computer science which one of them is machine learning and it play a key role on Machine Learning algorithms, data analysis and classification. And Machine Learning is used to enable computer programs to decide without being explicitly programmed which its application is most close with mathematic (mostly Linear Algebra) as graphing, predicating and regression.*

In this paper we look at what linear algebra and Machine Learning are and as the category of machine learning is closely related to statistics, we have a brief explanation of some statistical concepts. Then we shows the application of linear algebra especially matrix and vector in machine learning algorithms such as linear regression and support vector machine.

***Key Words*: Linear Algebra, Machine Learning, Statistics**

1. INTRODUCTION

Machine Learning is about discovering structures and patterns that exist in a set of 'things'. This is done using the language of mathematics, so we have to translate each 'thing' into numbers somehow. For example a single number can't sum up all the relevant facts about a thing very well; normally 'interesting' things are more complex. Instead we take some number of different measurements on each thing, and collect them into a vector of numbers that stands in for the thing itself.

Not just any mathematical genre but it is the linear algebra that is quite effective in regard to understanding Machine Learning technologies better. Linear algebra allows us to represent all the conditions based on the flexibility of vector space while allowing individuals to map data to specific dimensions, precisely within the vector space.

Experimental science, because of its empirical nature, uses mathematics as its primary tool, and scientists in the field are trying to help develop and extend these sciences by adapting their observations and experiences to mathematical principles. One of the empirical sciences is machine learning, a branch of artificial intelligence that has made extensive use of mathematics in its algorithms, models, structures, and predictions. In this paper we have used part of mathematics, namely linear algebra in machine learning, which is widely used in machine learning and even developers and engineers cannot develop the application without it, and discovering the Application of Linear Algebra in Machine Learning.

2. Introduction to Linear Algebra

At the core of linear algebra lies a very simple idea: linearity. A function f is linear if it obeys the equation $f(ax_1+bx_2)=af(x_1)+bf(x_2)$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are any two inputs suitable for the function. We use the term linear combination to describe any expression constructed from a set of variables by multiplying each variable by a constant and adding the results [3].

A significant proportion of the models used by scientists and engineers describe linear relationships between quantities. Scientists, engineers, statisticians, business folk, and politicians develop and use linear models to make sense of the systems they study. In fact, linear models are often used to model even nonlinear (more complicated) phenomena. There are several good reasons for using linear models. The first reason is that linear models are very good at approximating the real world. Linear models for nonlinear phenomena are referred to as linear approximations. The tangent line to a curve f(x) at xo is given by the equation $T(x)=f'(x_0)(x-x_0)+f(x_0)$.

This line has slope $f^0(x_o)$ and passes through the point $(x_o, f(x_o))$. The equation of the tangent line $T(x)$ serves to approximate the function $f(x)$ near x_o. Using linear algebra techniques to model nonlinear phenomena can be understood as a multivariable generalization of this idea.

Linear models can also be combined with nonlinear transformations of the model's inputs or outputs to describe nonlinear phenomena. These techniques are often employed 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 smoothly-varying output of a linear model into a hard yes or no decision [2].

Perhaps the main reason linear models are widely used is because they are easy to describe mathematically, and easy to "fit" to real-world systems. We can obtain the parameters of a linear model for a real-world system by analyzing its behavior for relatively few inputs.

Take for an Example at an art event, you enter a room with a multimedia setup. A drawing canvas on a tablet computer is projected on a giant screen. Anything you draw on the tablet will instantly appear projected on the giant screen. The user interface on the tablet screen doesn't give any indication about how to hold the tablet "right side up." What is the fastest way to find the correct orientation of the tablet so your drawing will not appear rotated or upside-down?

This situation is directly analogous to the tasks scientists face every day when trying to model real-world systems. The canvas on the tablet describes a two-dimensional input space, and the wall projection is a two-dimensional output space. We're looking for the unknown transformation T that maps the pixels of the tablet screen (the input space) to colored dots on the wall (the output space). If the unknown transformation T is a linear transformation, we can learn its parameters very quickly.

Let's describe each pixel in the input space with a pair of coordinates (x, y) and each point on the wall with another pair of coordinates $(\acute{x}, \acute{y})$. The unknown transformation T describes the mapping of pixel coordinates to wall coordinates:

$$(x, y) \xrightarrow{T} (x', y').$$

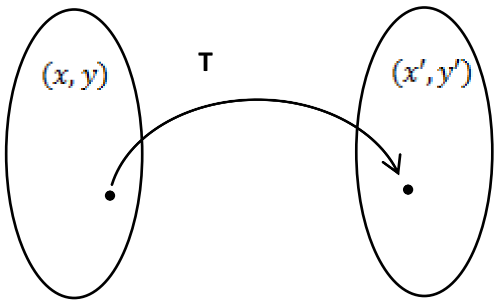

Fig 1: An unknown linear transformation T maps "tablet coordinates" to "screen coordinates."

To uncover how T transforms (x, y)-coordinates to $(\acute{x}, \acute{y})$ -coordintes, you can use the following three-step procedure. First put a dot in the lower left corner of the tablet to represent the origin $(0,0)$ of the $xy-coordinate\ system$. Observe the location where the dot appears on the wall—we'll call this location the origin of the $x\acute{y}'-coordinate$ system. Next, make a short horizontal swipe on the screen to represent the $x-direction\,(1,0)$ and observe the transformed $T\,(1,0)$ that appears on the wall. As the final step, make a vertical swipe in the y-direction $(0,1)$ and see the transformed $T\,(0,1)$ that appears on the wall. By noting how the $xy-coordinate$ system is mapped to the $x\acute{y}'$ coordinate system, you can determine which orientation you must hold the tablet for your drawing to appear upright when projected on the wall. Knowing the outputs of a linear transformation T for all "directions" in its inputs space allows us to completely characterize T.

In the case of the multimedia setup at the art event, we're looking for an unknown transformation T from a two-dimensional input space to a two-dimensional output space. Since T is a linear transformation, it's possible to completely describe T with only two swipes [2].

3. Introduction to Machine Learning

Machine learning is a data analysis method that automates analytical modeling. Machine learning is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention [11].

One of the definitions of machine learning as presented by Tom Michel Professor of Carnegie Mellon University is this: A computer program that, according to some T-tasks and P-function, forms the E experience, if its function in the T-task. Improve as measured by P with E experience [12].For example if a computer program can improve its performance in a task using its previous experiences then you can say that the machine has learned. And machine learning is the same as extracting knowledge from data [13].

In common computational methods, algorithms are written as a set of explicit commands to perform computation or problem solving for the purpose of running computers, but in machine learning the systems receive multiple inputs and perform statistical analysis to generate outputs in a single interval. Statistically, they themselves learn and so machine learning makes it possible for computers to automatically be able to perform the Decision Making process on new input data after receiving the sample data and modeling it.

3.1 Types of Machine Learning

Here are the types of machine learning with great related examples.

3.1.1 Supervised Learning

In order to train the algorithm in supervised learning, data is provided as a primary input to the system previously labeled by the human factor with the desired and expected outputs and hence the inputs and outputs corresponding to them as pairs of input /Outputs are provided to the system [14].

The purpose of presenting these input / output pairs is to identify the existing errors by comparing the correct predetermined response and the predicted response, thereby generating its own trained outputs and the model used. Modify based on these trainings and the training process will continue until the prediction model of the algorithm is sufficiently accurate and the prediction results are consistent with the predicted result.

And so in post-training supervised learning the system will be able to predict the appropriate output in the face of the so-called Unlabeled data (including supervised learning algorithms can be regression, Decision Tree, Random Forest, KNN and Logistic Regression) [15].

One of the most common uses of supervised learning is when past information is used to predict future events. For example, this method can be used to forecast stock market fluctuations in the coming weeks and months in the last few months or weeks. Another example of using this algorithm is to detect spam emails from non-spammers.

3.1.2 Unsupervised Learning

In unsupervised learning, the input data lacks any tags, so the learning algorithm looks for common similarities and features of the input data, and since unlabeled data is more common and pervasive than labeled data, this learning method is particularly important. In other words, the goal of unsupervised learning is to identify hidden patterns in unlabeled data sets [16]. In fact, by using unsupervised learning algorithms, the system will be able to identify different data properties and use them to categorize the data considered.

3.1.3 Reinforcement Learning

In reinforcement learning, the algorithm learns to make certain decisions by going through a trial-and-error cycle, thereby refining its predictions and subsequent decisions and continuously learning them according to previous decisions (Markov Decision Process is one of the reinforcement learning algorithms) [17].

To clarify, and as a practical example, we can consider a footballing robot that, by positioning itself in different situations and making decisions appropriate to these situations and gradually correcting its errors, eventually learns that in each situation to take the most correct decision to shoot.

3.2 Machine Learning and Statistics

Statistics is the study of collecting, analyzing, and interpreting data with mathematical rules and ultimately the application of their results. As the category of machine learning is closely related to statistics, here are a brief explanation of some statistical concepts:

3.2.1 Correlation

Correlation is the degree of dependence of two variables on each other, which is represented by a coefficient called the correlation coefficient whose value is a number between 0 and -1. The correlation coefficient of 0 between the two parameters means that these parameters have no relation to each other And as long as the correlation coefficient of the two parameters is farther than 0, This means that the changes in the two parameters are more interdependent [18]. The positive correlation coefficient also means that if one of the two parameters increases, the other one will increase and if one of them decreases, the other one will decrease, but the negative correlation coefficient means the inverse correlation between the two parameters. In other words, if one decreases, the other increases, and if one increases, the other decreases. For example, the correlation between human height and weight is usually a positive correlation, and the taller people are, the higher their weight (although there will always be exceptions).

3.2.2 Regression

Regression literally means "return". When the two variables are highly correlated, the regression makes it possible to predict and express changes in one variable based on changes in another.

3.2.3 Mean

Sometimes known as the Arithmetic Mean or Average Score, is a type of statistical examination of the central obtained from the sum of the values in a dataset divided by their number. (The tendency for the center refers to numbers that summarize pre-measured data.) The mathematical formula for calculating the mean is as follows [21]:

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

3.2.4 Median

In statistics, the mean of the mean, denoted by M, is a survey of the degree of tendency toward the center, which divides the statistical population into two equal parts, which, if the number of members is odd, will be considered the number in the middle. However, if the number of members of the statistical population is even, the average is obtained from the average of the two members in the middle of the statistical population (Need to explain that to calculate the median, you first need to sort the data from small to large and then select the number that is in the middle as the median) [21].

3.2.5 Mode

Mode is a type of centrally-oriented survey that refers to the data most frequently repeated in a set of data in statistics. Unlike the mean and the median we only have one number, as we said, in a statistical society we can have more than one mode or no mode.

3.2.6 Variance and Standard Deviation

In statistics, Variance is a measure of dispersion. In other words, it is a numerical variance that shows how the data is distributed around the mean value calculated by the following mathematical formula:

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

In other words, to calculate the variance, we first obtain the mean, as previously described, then subtract each member of the statistical population from the mean and bring the result to power.

3.2.7 Standard Deviation

Standard Deviation is also one of the indicators of data dispersion in statistics that is obtained from the variance root, showing how much the average data is far from the mean. Note that if the standard deviation of a set of data is close to zero, this indicates that the data are close to average and have low dispersion, while the larger standard deviation indicates a significant dispersion of data [22]. The mathematical formula for calculating the standard deviation is as follows:

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

Keep in mind that standard deviation provides valuable data so that by using the SD value we find out which members of the statistical community are within the standard range, which members are above the standard limit and which one is below the standard limit.

4. Introduction to Machine Learning Algorithms

The algorithms and approaches used in machine learning are constantly changing and evolving, with some of the most common ones discussed below.

4.1 Linear regression algorithm

Linear regression is one of the most commonly used methods of data modeling that has a very simple mathematical basis. Whenever we can identify a linear relationship between two variables, we can use this kind of regression to predict the values of these variables based on the value of the other variable [32].

The purpose of the linear relationship is to see that by increasing one variable, the other variable increases (or decreases) And by decreasing it, the second variable mutually decreases (or increases), and this increase or decrease has a direct relation (simple coefficient) to the value of the first variable, which we call the independent variable.

An easy way to figure out this relationship is to plot the value of one variable on the other in a graph, and if the resulting figure was like a straight line, we can conclude that the relationship between these two variables is a linear relationship. This graph is also called the distribution graph.

The following graph simply displays the height and weight values of the twenty people obtained from the actual data in a distribution graph.

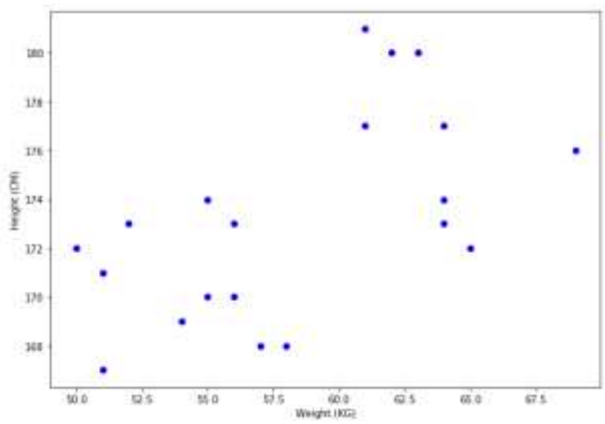

Chart-1: height and weight values of the twenty people.

After finding the linear relationship between the two variables, we simply calculate the linear relationship formula between them and use it to predict new values of one variable based on the value of the other variable. That's exactly what linear regression does, finding the linear relationship between two variables by calculating the coefficients of a linear relationship and using it to predict values.

The formula for a line is: Y = w1 * X + w0 where w1 is the slope of the line and w0 is the width of the source. In this relation, X is called the independent variable and Y is the dependent variable, although it is easy to obtain X in Y and replace the dependent and independent variable as desired.

4.1.1 Mathematical calculations of linear equation coefficients

To find the formula of a line, we just have two points, but in the real world, instead of two points, we have thousands of data to find the best fit line. So first we need to find the definition of "best" and then find the coefficients of the line equation.

We want to find a relationship between the area of a house in square feet and its price in dollars, and we know that there is a linear relationship between the two. (The dataset from this address is viewable and downloadable [34]). To do this, we need to find the best line that shows the price relationship by area of the house. Suppose we find this line, that is, for every house with a known area, its approximate price is obtained by putting in this relation.

Now, if we include the available data (selling homes, also called training data) in this regard, for each house, an estimated selling price will be obtained which will be different from the actual home price. Our goal in linear regression is to find a line that has the least difference between the available data and the estimated data.

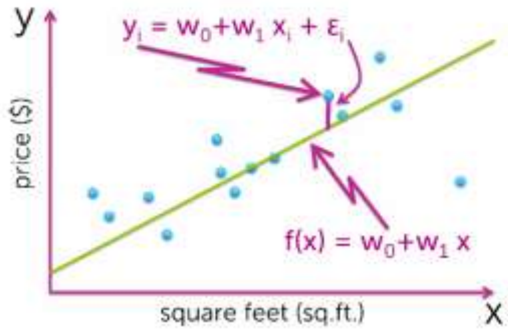

Chart -2: Relationship between area of the house and their prices.

So if we have a function that shows the difference between the real values and the estimated values, such as the sum of the squares of these differences (if we add up the differences ourselves because we have positive and negative numbers, they will neutralize each other and produce an irrational number.), We can look for a way to find two coefficients W0 and W1 that are the same line coefficients to minimize this function. Also called the Residual Sum of Squares

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

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

here,

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

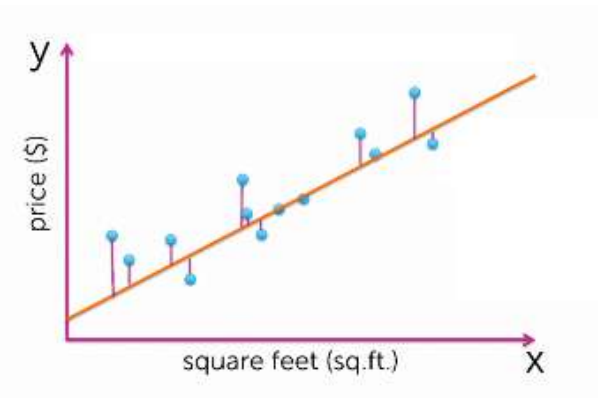

Chart -3: Residual sum of squares (RSS) diagram.

RSS (W0, W1) is a quadratic function and for different values of W0 and W1, having x and y, it produces different outputs as you see in the following figure in three-dimensional form:

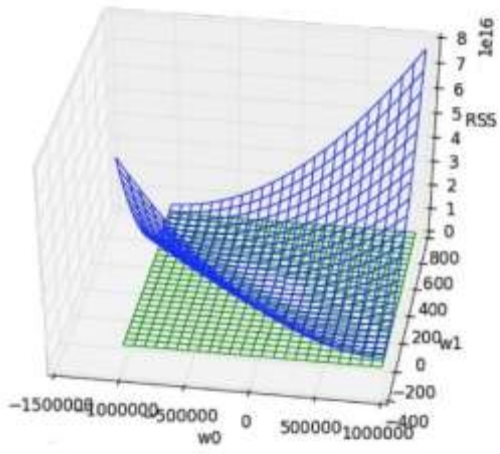

Chart -4: 3D plot of RSS with tangent plane at minimum [35].

Minimize function over all possible w0, and w1 $\sum_{i=1}^{N} (y - [w_0 + w_1 x_i])^2$ which $RSS(w_0, w_1)$ is a function of 2 variables.

Mathematical arguments to minimize this function are beyond the reach of this discussion but looking at the figure above, we see that first there is this global minimum and secondly this minimum occurs when That the RSS changes to W0 and W1 are zero (To visualize 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 this function is taken from both variables and set to zero (although in the real world the gradient reduction method is used). The result is:

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

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

Where N is the number of available data, x and y are both problem data and $\bar{x}$ is the mean of our independent variable. If you follow the formula above, W1 or the slope of the line is the correlation coefficient between the two variables x and y, and W0 is the mean difference of y and the multiplication correlation coefficient of x.

4.1.2 Interpretation of Linear Regression Coefficients

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 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 W1 interpretation. w1 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.

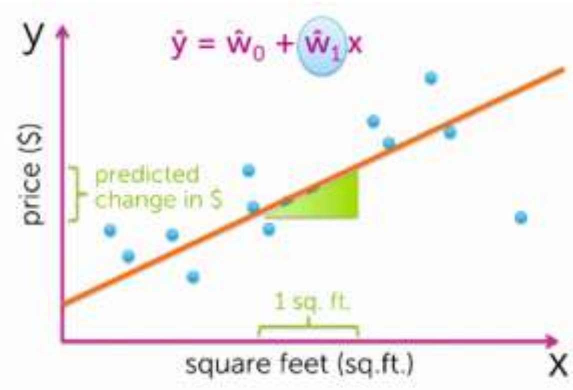

Chart -5: Diagram with little different slop [36].

4.1.3 Multivariate regression

Multilinear regression is a generalization of linear regression considering more than one independent variable and a particular case of general linear models formed by limiting the number of variables dependent on a fundamental model for linear regression that is:

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

In the above relation we assume that n observations of a dependent variable and p are independent variables. Then i is $y_i th$ observation of the dependent variable, x_{ij} also is $i\ th$ observation of the j dependent variable which p is j=1, 2, 3… amounts of β_j Indicates the estimated variables and ε_i also is the $i\ th$ distributed independent normal error.

For example, suppose we need a combination of weight and blood fat to predict a person's blood pressure. We need to use multivariate regression here.

Note: in multivariate regression there should be two important assumptions, first that the number of independent variables is not greater than the number of observations and data available, which is usually always the case, and second, that there is no significant linear relationship between the independent variables themselves. For example, if there is a linear relationship between weight and blood fat, we cannot use the linear combination of these to predict blood pressure.

4.1.4 Four main assumptions in the use of linear regression

To use this simple and efficient technique in predicting the values of a dependent variable, the following four conditions must be met in order to account for the validity of the proposed model (data discovery relation).

The data should be normally distributed: That is, we should not divert data as much as possible and our statistical community should be a true example of the whole community. On the other hand, if we have a lot of divergent data, we first delete them and create a uniform dataset.

A linear relationship between independent variables and dependent variables: Our main assumption in linear regression is that there is a linear relationship between variables that we must be sure of, or else our modeling is unreliable. One way to mathematically investigate this issue (other than graphing the distribution and visualizing the linear relationship) is to draw an RSS graph, or the power of two estimated values of actual values based on the dependent variables. The diagram should be perfectly flat and linear (diagram to the right of the figure below). Otherwise, the linear relationship between the dependent and independent variables is not guaranteed.

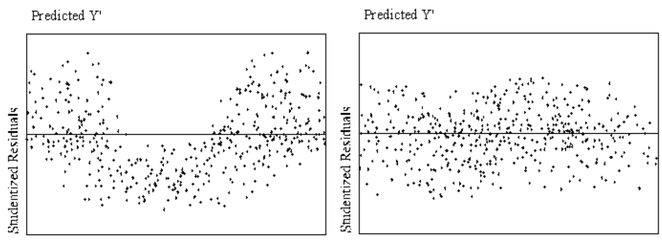

Fig -2: Linear relationship between independent and independent variable.

Variables and data should be carefully collected: Otherwise the model created will not fit the real world.

Homoscedasticity: The variance of the error between all independent variable values should be normally distributed and the shape of its distribution graph illustrates this.

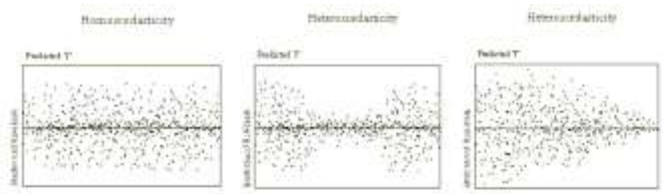

Fig -3: Homoscedasticity presentation.

4.2 Support Vector Machine Algorithm

SVM (support vector machine) is one of the most common algorithms and methods in the field of data classification. Suppose we have a dataset where 50% of the population is male and 50% is female. This dataset can be customers of an online store. By having a subset of this data that identifies the gender of individuals, we want to create rules that will help us determine the gender of the rest of the population with great accuracy. Detecting the gender of store visitors allows us to display separate ads for men and women and increase store profitability. We call this process classification in data science.

To fully explain the problem, suppose two of the parameters we are going to determine are gender, height, and hair length. The graph of the height and length distribution of individuals is shown below, where the sexes of individuals with two square (male) and circle (female) symbols are shown separately.

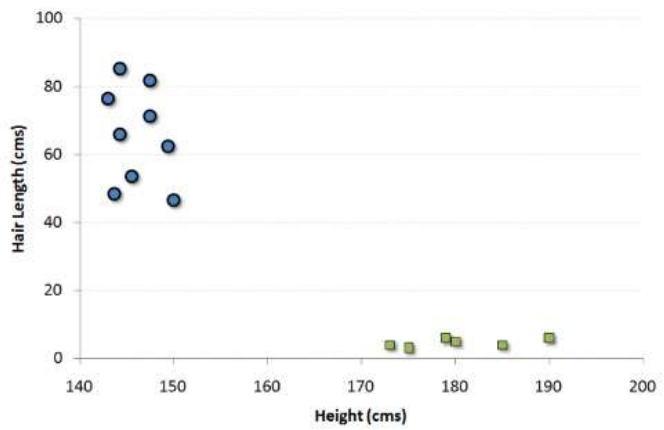

Chart -6: Height and hair length graph [37].

Looking at the graph above, the following facts are easily visible:

1. Men in this collection have a higher average height.
2. Women have longer average hair lengths.

If we were given a new data with 180cm height and 4cm hair length, our best guess for this person would be the men's category.

4.2.1 What is a Support Vector and what is SVM?

In simple language, backup vectors are a set of points in the n-dimensional space of the data that defines the boundaries of the categories, and the bounding and categorization of the data is done by them, and by moving one of them, the output of the categorization may change. For example, in the above figure, the vector (45,150) is a member of a support vector belonging to a woman. In 2D space, the support vector will form a line, in a 3D space form a page and in an n-dimensional space form a cloud.

An SVM is a cluster or boundary that measures the best categorization and separation of data by benchmarking support vectors.

In SVM, only the data contained in the support vectors are based on machine learning and model building, and this algorithm is not sensitive to other data points and aims to find the best boundary between the data so that it is as far as possible from all categories (their support vectors).

4.2.2 How to create a machine based on support vectors?

In addition to the data in the example above, we can have many boundaries, three of which are shown below.

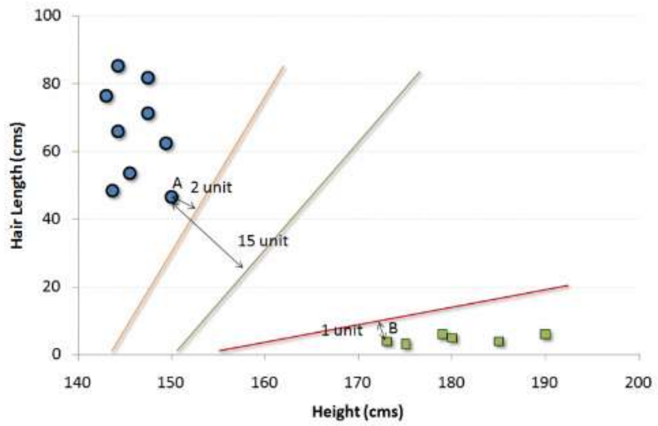

Chart -7: The 13 figure with three boundaries [37].

The question is, what is the best line in this issue?

An easy way to do this is to create an optimal cluster, calculate the distance of the boundaries obtained with the support vectors of each cluster (the most boundary points of each cluster or class), and finally select the boundary that is the largest of all available clusters. Which in the form of the midline above, a good approximation of the border that is far from both.

4.2.3 Nonlinear data distribution and application of support vector machine

If the data are linearly separable, the above algorithm can create the best machine for separating data and classifying a data record, but how to determine SVM if the data is not linearly distributed (as shown below).

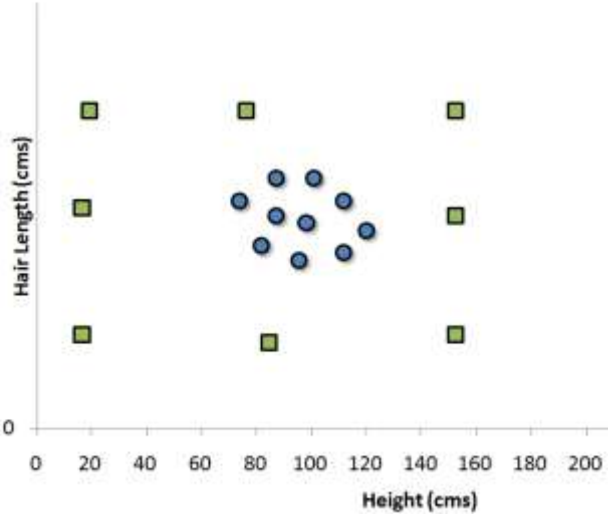

Chart -8: Linearly data distribution graph [37].

In this case, we need to map the data to another space using a kernel function where the data is separable and SVM can be easily defined. Correct determination of this mapping function is effective in the performance of the support vector machine. Assuming the conversion function for the above example, our data space will be converted to:

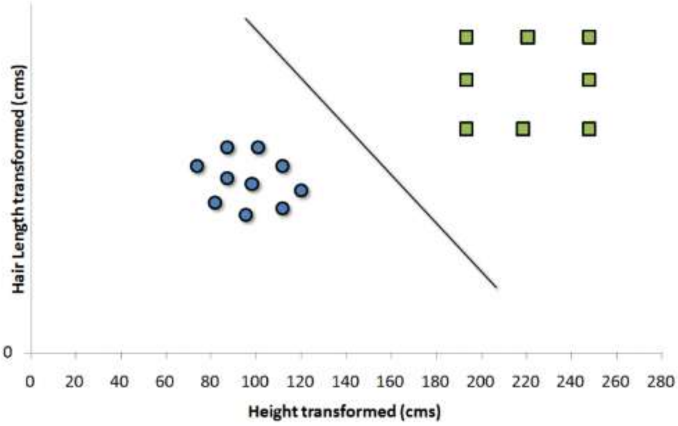

Chart -9: Mapping data using kernel function [37].

In this transformed space, it's easy to find an SVM.

4.2.4 A closer look at the SVM process

As mentioned, SVM moves data to a new space according to their predefined batches So that data can be classified and segmented linearly (or cloud page) And then by finding the supported lines (pages supported in multidimensional space), tries to find the linear equation that creates the greatest distance between the two categories.

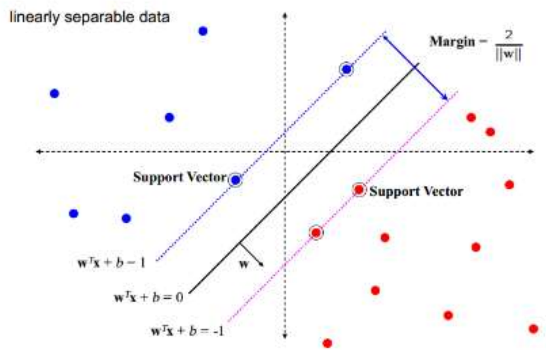

Chart -10: SVM diagram with good representative [38].

In the figure above, the data are shown in red and blue, respectively, and the dotted lines represent the support vectors corresponding to each category, denoted by double circles, and the continuous black line is the SVM. The supported vectors each have a characteristic formula that describes the boundary line of each category.

In Python the three SVC, NuSVC, LinearSVC functions are the main task of the classification. An example of a category with these functions can be found below [39]:

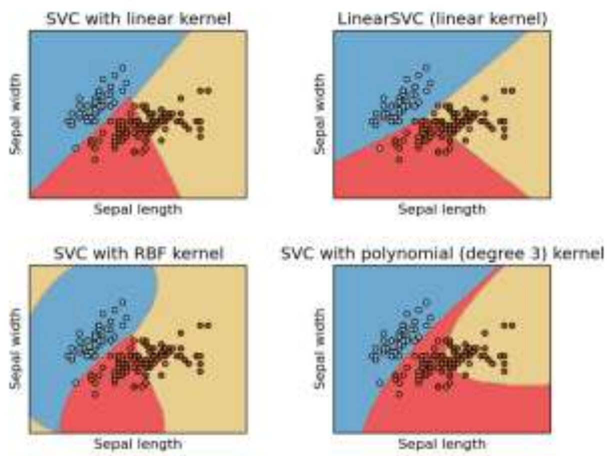

Fig -4: SVM with four function [40].

You can also use this article for a more accurate selection of the kernel type and its parameters [41].

4.2.5 Support vector machine in action

To use SVM for real data, you need to follow several tips to get acceptable results.

1. First, filter the data (Outliers or Missing data)

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

2. Numeracies and normalize data. In short, convert data such as gender, field of study and so on to numbers and try to normalize the values of all the attributes between one and minus one [1, -1] so that large or small values of a data attribute do not affect the machine.

3. Try different kernels and, in addition to each, depending on the training dataset you have and the classification of their data, measure the accuracy of the SVM and if necessary change the parameters of the conversion functions to get the better answers. Try this work for different kernels as well. You can start from the RBF kernel.

4.2.6 Weaknesses of support vector machine

- These types of algorithms have inherent limitations, for example, it is not yet clear how to set parameters by attaching a mapping function.
- Support vector based machines require complex and time consuming computations and consume a lot of memory due to their complexity.
- Discrete and non-numeric data are also incompatible with this method and must be converted.

However, SVMs have a coherent theoretical foundation and the solutions produced by them are global and unique. Today, support vector machines have become the most common predictive techniques in data mining.

5. CONCLUSIONS

In this paper we introduced to Linear Algebra, Machine Learning and Machine Learning algorithms such as linear regression and support vector machine that explicitly have linear relationships to independent and non-independent variables, using metrics for data-sets and the vector is used for classification and boundaries of datasets in n-dimensional space.

Linear algebra is something Machine Learning experts can't live without. You will never evolve from amateur to good specialist without mastering this field of applied mathematics. Usually math is not the key to success, but in many cases, it helps, and in a few (like developing deep learning models), it is essential. Linear algebra ideas and methods are easiest to comprehend while solving interesting problems, it will help to understand abstract concepts the best. You can see that the deeper you dive into machine learning, the more linear algebra you see there.

Linear regression is an old method from statistics for describing the relationships between variables. It is often 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 [37].

REFERENCES

[1] jupyter.readthedocs, "Installing Jupyter Notebook," 6 10 2019. [Online]. Available: https://jupyter.readthedocs.io/en/latest/install.html.

[2] I. Savov, No bullshit guide to linear algebra, Montr�l, Qu�ec, Canada: Minirefer ence Co, 2016.

[3] O. Kharkovyna, "Mathematics for AI: Linear Algebra and How to Understand It Better," 12 10 2019. [Online]. Available: https://towardsdatascience.com/mathematics-for-ai-linear-algebra-and-how-to-understand-it-better-63b430999069.

[10] R. SOBERS, "Artificial Intelligence vs. Machine Learning in Cybersecurity," 9 9 2019. [Online]. Available: https://www.varonis.com/blog/ai-vs-ml-in-cybersecurity/. [Accessed 13 10 2019].

[11] sas, "machine learning," 7 11 2019. [Online]. Available: https://www.sas.com/en_us/insights/analytics/machine-learning.html.

[12] D. Dwivedi, "Machine Learning For Beginners," towardsdatascience., 7 5 2018. [Online]. Available: https://towardsdatascience.com/machine-learning-for-beginners-d247a9420dab. [Accessed 25 9 2019].

[13] M. Mosawi, "what is machine learning," digiato, 13 7 2015. [Online]. Available: https://digiato.com/article/2015/07/13/. [Accessed 28 9 2019].

[14] J. Brownlee, "Supervised and Unsupervised Machine Learning Algorithms," 12 8 2019. [Online]. Available: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/.

[15] J. Brownlee, "A Tour of Machine Learning Algorithms," 5 12 2019. [Online]. Available: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/.

[16] S. Mishra, "Unsupervised Learning and Data Clustering," 20 5 2017. [Online]. Available: https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a.

[17] geeksforgeeks, "Reinforcement learning," [Online]. Available: https://www.geeksforgeeks.org/what-is-reinforcement-learning/. [Accessed 13 9 2019].

[18] statisticssolutions, "Correlation (Pearson, Kendall, Spearman)," [Online]. Available: https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/. [Accessed 9 10 2019].

[19] W. KENTON, "Descriptive Statistics," 27 6 2019. [Online]. Available: https://www.investopedia.com/terms/d/descriptive_statistics.asp.

[20] laerd, "Descriptive and Inferential Statistics," 7 11 2019. [Online]. Available: https://statistics.laerd.com/statistical-guides/descriptive-inferential-statistics.php.

[21] analyzemath, "Mean, Median and Mode of a Data Set," [Online]. Available: https://www.analyzemath.com/statistics/mean-median-mode.html. [Accessed 18 10 2019].

[22] wikipedia, "Standard deviation," [Online]. Available: https://en.wikipedia.org/wiki/Standard_deviation. [Accessed 1 12 2019].

[23] numpy.org, "numpy," [Online]. Available: https://numpy.org/. [Accessed 12 11 2019].

[24] anaconda, "Anaconda Documentation," [Online]. Available: https://docs.anaconda.com/. [Accessed 5 11 2019].

[25] M. D. Okoi, "Popular Programming Languages on GitHub," fossmint, 30 4 2019. [Online]. Available: https://www.fossmint.com/popular-programming-languages-on-github/. [Accessed 1 10 2019].

[26] N. Heath, "top 10 programming languages for machine learning," techrepublic, 25 2 2019. [Online]. Available: https://www.techrepublic.com/article/github-the-top-10-programming-languages-for-machine-learning/. [Accessed 2 10 2019].

[27] Kite, "TensorFlow or PyTorch? A guide to Python machine learning libraries," 26 9 2019. [Online]. Available: https://aibusiness.com/tensorflow-or-pytorch-a-guide-to-python-machine-learning-libraries/.

[28] renjie-liu, "tensorflow," github, 2 9 2019. [Online]. Available: https://github.com/tensorflow/tensorflow. [Accessed 5 10 2019].

[29] F. Copes, "The Complete JavaScript Handbook," freecodecamp, 30 10 2018. [Online]. Available: https://www.freecodecamp.org/news/the-complete-javascript-handbook-f26b2c71719c/. [Accessed 6 10 2019].

[30] dev, "The best programming language for Artificial Intelligence and Machine Learning," [Online]. Available: https://dev.to/duomly/the-best-programming-language-for-artificial-intelligence-and-machine-learning-e7b. [Accessed 17 11 2019].

[31] julialang.org, "Julia 1.0," julialang, 8 8 2018. [Online]. Available: https://julialang.org/blog/2018/08/one-point-zero. [Accessed 6 10 2019].

[32] statisticssolutions, "What is Linear Regression?," [Online]. Available: https://www.statisticssolutions.com/what-is-linear-regression/. [Accessed 10 11 2019].

[33] wikipedia, "Regression analysis," [Online]. Available: https://en.wikipedia.org/wiki/Regression_analysis. [Accessed 26 10 2019].

[34] [Online].Available: https://github.com/Jonasyao/Machine-Learning-Specialization-University-of-Washington-/blob/master/Regression/Assignment_four/kc_house_data.csv .[Accessed 7 10 2019]

[35] arafatm, "edu_coursera_machine_learning_1_foundations," 28 5 2017. [Online]. Available: https://github.com/arafatm/edu_coursera_machine_learning_1_foundations/blob/master/README.md.

[36] bigdata, "2016 5 13 ",رگرسیون-خطی-و-یادگیری-ماشین. [Online]. Available: http://www.bigdata.ir/1395/02/رگرسیون-خطی-و-یادگیری-ماشین.

[37] analyticsvidhya, "support vector machine simplified," 10 2014. [Online]. Available: https://www.analyticsvidhya.com/blog/2014/10/support-vector-machine-simplified/.

[38] semanticscholar, "Simulation of Groundwater by using Support Vector," [Online]. Available: https://www.semanticscholar.org/paper/Simulation-of-Groundwater-by-using-Support-Vector-(-Plain-Mehrdad/2c70a1462bb52d872f814aa8883dec4cfaa32ebc. [Accessed 4 10 2019].

[39] scikit-learn, "sklearn svm," [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html. [Accessed 6 10 2019].

[40] maviator, "SVC," 5 1 2018. [Online]. Available: https://maviator.github.io/2018/01/05/SVC/.

[41] csie.ntu.edu, support vector machine guide, .csie.ntu.edu.tw.[Accessed 8 10 2019]

[42] [Online].Available:http://www.bigdata.ir/wpcontent/plugins/downloadattachments/includes/download.php?id=1646 .[Accessed 5 10 2019]