International Journal of Engineering and Techniques-Volume 8 Issue 2, March - April - 2022

A Model Based Linear Algebraic Approach for Machine Learning

[1]Shaik Kaleshavali, [2]Dr. Ashish Verma, [3]Dr. MD. Nizam

[1]Research Scholar, Department of Mathematics , professor,SCET, Peerancheruvu, Hyderabad, [2]Professor, Department of Mathematics, V.B.S Purvanchal University, Jaipur, India [3]Associate Professor,, Department of Mathematics, Nalla Malla Reddy Engineering college, Hyderabad, India.

E-Mail: kaleshavalishaik2555@gmail.com, nizam.mps@nmrec.edu.in

Abstract

Linear algebra is sub field of mathematics and contains matrix's, Operations on dataset, Vectors. Linear algebra is core base for purely statistics and mathematics person to achieve goals with basic logics of linear algebra using machine learning. As a machine learning aspirant, the aspirant should be best at linear algebra to work on creating a new module or bundle. In other areas, such as statistics, engineering, research, linear algebra plays an important role and has strongest impact Although solving any problem mathematically with linear algebra is crucial for experienced engineers, linear algebra is crucial in machine learning, and in the case of any practical implementation of linear algebra, it is quite probable that the technical person may not have mathematical knowledge. Easier implementation is provided by linear algebra. Coronavirus disease (COVID-19) has raised urgent questions regarding minimizing and improving effective analytical strategies for analyzing information obtained from standardized protocols. In order to understand what quality of care interventions are the most successful and to test them as quickly as possible, it becomes necessary to evaluate available data immediately. It is necessary to integrate and clean up data from large multi-center hospitals and needs sophisticated data processing. To collect and store their data in a standardized manner, case report forms (CRF) for patients with suspected or confirmed COVID-19 are required. The goal of this study is to determine if it is possible to avoid or postpone local outbreaks of COVID-19 by travel restrictions from abroad. The problem statement is that we

are unable to gather clear suspicious data from big data of covid-19 patients, and during the corona outbreak it becomes difficult to distinguish symptotic and asymptotic patients. The number of beds required or ventilators needed for COVID-19 cases at this point is difficult to estimate.

Keywords—Dimension Reduction, Singular Value Decomposition, Singular Vectors & Singular Values, Eigen Value and Eigen Vector.

INTRODUCTION

Linear Equation is an equation which may be represented as follows

$$a_1x_1 + \cdots + a_nx_n + b = 0,$$

Where X1……..Xn are variables, b, a1……an are the coefficients which are often real numbers

Linear algebra is a part of mathematics with respect to linear equations

$$a_1x_1 + \cdots + a_nx_n = b,$$

Linear mapping as,

$$(x_1, \ldots, x_n) \mapsto a_1x_1 + \ldots + a_nx_n,$$

Consists of vector spaces and through matrices

The reason to use Linear algebra, it is sub field of mathematics and contains matrix's, Operations on dataset, Vectors. Linear algebra is core base for purely statistics and mathematics person to achieve goals with basic logics of linear algebra using machine learning.

As a Machine learning aspirant to work on developing new module or package the aspirant should be best at linear algebra. Linear algebra plays an important role and have best impact on other fields such as Statistics, Engineering, and Science.

The reason why linear algebra essential in machine learning, while solving any problem mathematically with linear algebra is crucial for technical person and in case of some practical implementation of linear algebra it is quite possible the technical person won't have mathematical knowledge person is good at his programming skills. Linear algebra supports easier implementation.

Linear Algebra's Field to be studied

5 fields of linear algebra every Machine Learning aspirant must learn are

- Linear Algebra Notation
- Linear Algebra Arithmetic
- Linear Algebra for Statistics
- Factorization of matrix

Linear Algebra Notation: Learn to write and read matrix and vector. Notation allows specific operators to accurately diagnose operations on data.

Linear Algebra Arithmetic: Arithmetic operations are performed by pairing with linear algebra notation. Arithmetic operations consist of scalars, vectors and matrices being added, subtracted, and multiplied.

There are several difficulties facing freshers in the field of linear algebra for performing operations such as matrix multiplication, tensor multiplication that are not implemented as direct element multiplication.

Linear Algebra for statistics: Describing, understanding, sampling data is concerned to statistics. Chunks of data is analysed and then data cleaning process takes place in machine learning.

Factorization of Matrix: Constructing on arithmetic and notation is the core concept of matric factorization also known as matrix decomposition. In linear algebra, factoring matrix is an enhance the integrity. Elements are commonly used in many more complex operations in both linear algebra (inverse matrix) and machine learning (Principal Component Analysis, SingularValue Decomposition, least square).

II. MACHINE LEARNING TECHNIQUES WITH LINEAR ALGEBRA

Linear algebra is a short notation for describing some compact operation. Some common machine learning technique via linear algebra are describe below

1. Principal Component Analysis (PCA)

2. Linear Least Square for linear Regression

3. Singular Value Decomposition (SVD)

A.Principal Component Analysis (PCA)Data Compression: Data compression is a method of reducing the number of bits needed for either data storage or transmission. Data compression is categorized into methods that are either lossy or lossless. In lossy technique, as name suggests there is loss of information or data. Where there is no loss, as in a lossless method. There are various forms of algorithm for coding, such as Huffman coding, coding for run length.

Lossy Compression technique

Lossy Compression technique at the cost of data quality one can achieve higher compression ratio. There are two methods of lossy data compression:

- Wavelet transforms

- Principal Component Analysis PCA

Principal Component Analysis (PCA): A representation of data with orthogonal fundamental vectors is provided by Principal Component Analysis (PCA). Eigen values of data matrix covariance. This original projection dataset is reduced with little loss of knowledge, which is derived using the singular value decomposition method. PCA is interpreted by the covariance matrix's own value/eigen vector approach. Before moving forward let's understand core concepts like Dimensionality, correlation, orthogonal, eigen vectors, covariance matrix

<u>Dimensionality:</u> The number of random variables in a dataset or either the number of features or, more specifically, the number of columns in the dataset.

<u>Correlation:</u> It shows how strongly two variables are correlated to each other. The same range value for +1 to-1. Positive indicates that as one variable increases, the other also increases. Whereas negative suggests the other decrease of the increase of the former.

<u>Orthogonal:</u> irrelevant to each other, ie. Correlation between any variable pair ie 0.

<u>Eigen vectors:</u> Eigen values and Eigen vectors are a major domain of themselves. Let's limit ourselves to the awareness of the same thing that we would need here. Consider non zero vector V. It is an eigen vector of square matrix A, if AV is scalar multiple of V.

$AV = \lambda . V$

V is eigen vector

λ is eigen value.

<u>Covariance matrix:</u> Layer is calculated consisting of covariances between pairs of variables. The (i, j) component is the covariance between the ith and the jth variable.

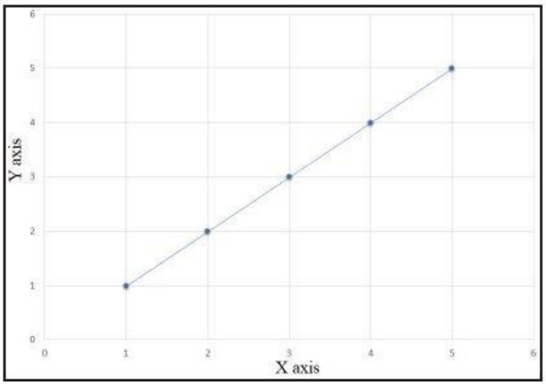
III. LINEAR LEAST SQUARE FOR LINEAR REGRESSION

It is also known as linear model; Linear regression is a predictive algorithm which mostly provides a linear relationship between both the [calls Y] prediction and the [call is X] data. If we plot an X, the Y linear relationship will always come up with a straight line. Example if we plot the graph of these values:

X= 1, 2,3,4,5 …… (input)

Y= 1, 2,3,4,5 …… (Output)

It will be perfect straight line.



Least square regression

The least square regression is a statistical technique that minimizes the error in such a way that the amount of all square error is minimized.

Steps to estimate the minimum square regression. Formula for the estimation of m=slope

$$m = \frac{\text{Sum of all } (X - xmean)*(Y - ymean)}{\text{Sum of } (X - xmean)**2}$$

**2 syntax in python Mean value of all 'x'= (1+2+3+4+5)/5 = 3 Calculating X – xmean for all X value

At X = 1: (X - xmean) = 1 - 3 = -2

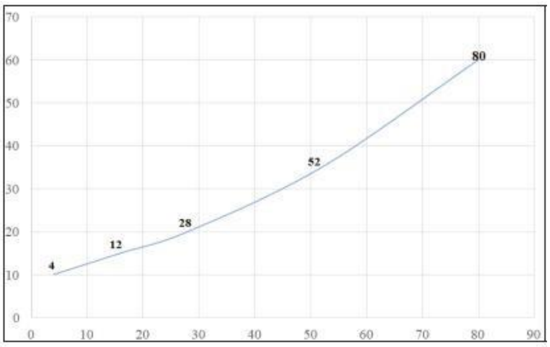
At X = 2: (X - xmean) = 2 - 3 = -1

At X = 3: (X - xmean) = 3 - 3 = 0

At X = 4: (X - xmean) = 4 - 3 = 1

At X = 5: (X - xmean) = 5 - 3 = 2

Calculating all Y – ymean values for all Y



Y – ymean is: (4+1+28+52+80)/5 = 35.2

At Y= 4: (Y - Ymean) = 4- 35.2= -31.2

At Y= 12: (Y - Ymean) = 12- 35.2= -23.2

At Y= 28: (Y - Ymean) = 28 - 35.2= -7.2

At Y= 52: (Y - Ymean) = 52 - 35.2= 16.8

At Y= 80: (Y - Ymean) = 80 - 35.2= 44.8

The availability of these values enables us to determine the sum of all values

(x,y) = (X-xmean)*(Y-ymean)

(1,4) = (X-xmean) *(Y-ymean) = (-2*-31.2) =62.4

(2,12) = (X-xmean) *(Y-ymean) = (-1*-23.2) =23.2

(3,28) = (X-xmean) *(Y-ymean) = (0*-7.2) =0 (4,52) = (X-xmean) *(Y-ymean) = (1*16.8) =16.8 (5,80) = (X-xmean) *(Y-ymean) = (2*44.8) =89.6 Sum of all = (62.4+23.2+0+16.8+89.6) =192 Now denominator part, Sum (X-xmean) **2 = sum (-2**2, -**2,0**2,1**2,2**2) =sum (4,1,0,1,4) =10

$$m = \frac{\text{Sum of all } (X - \text{xmean})*(Y - \text{ymean})}{\text{Sum of } (X - \text{xmean})**2}$$

=192/10=19.2

Calculate of Y intercept, formula

b= ymean – m * xmean =35.2-19.2*3 =35.2-57.6.

= -22.4

Overall formula can now be written inform of

Y= mx+b

Y= mx+b = 19.2x+ (-22.4)

Least Square regression

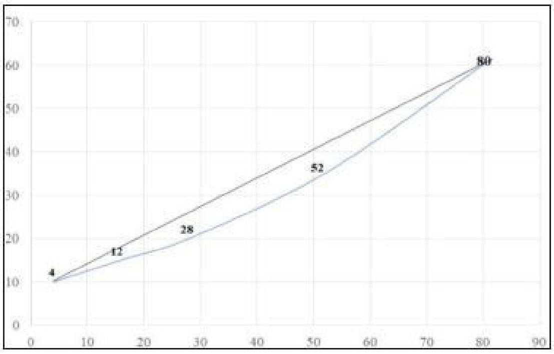
x=1→ y= 19.2 *1 – 22.4= -3.2

x=2 →y=19.2 * 2 – 22.4 = 16

x=3→ y= 19.2 * 3 – 22.4 = 35.2

x=4→ y= 19.2 * 4 – 22.4 = 54.4

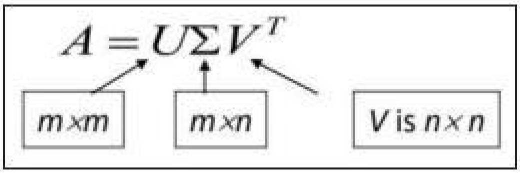
x=5→ y= 19.2 * 5 – 22.4 = 73. 6



IV. SINGULAR VALUE DECOMPOSITION

The solution for PCA in terms of the covariance matrix's Eigen vectors. There is, however, another way to obtain a solution based on a singular decomposition value, or SVD. This generally generalizes the notion of Eigen vectors from square matrices to any kind of matrix.

There is a factorization (Singular Value Decomposition = SVD) as follows for Am x n having a rank r:



Columns U are orthogonal eigenvectors of AA^T.

Columns V are orthogonal eigenvectors of $A^T A$.

Eigenvalues × 1 … × r of AA^T are the eigenvalues of $A^T A$.

$\sigma_i = \sqrt{\lambda_i}$ …...............Singular value

$$\Sigma = diag(\sigma_1 \ldots \sigma_r)$$ Singular vectors and singular values Properties of AA^T and A^TA:

- Symmetrical
- Square
- At least positive semi definite (eigen values are zero or positive)

- Both matrices have same positive eigen values Both have same rank r as A.

CONCLUSION

Extraction of related summaries easily and effectively. Reduces the complexity of space. High dimensional data visualization. Appropriate numbers for health planning can be created. Infected and stable patients may be identified during the healthy patient during the corona epidemic. ABBRIVATIONS PCA- Principal Component Analysis, SVD- Singular Value Decomposition.

REFERENCES

[1] EthemAlpaydın, "Introduction to Machine Learning Second edition", The MIT Press Cambridge, Massachusetts London, England.

[2] Kelvin P. Murphy, "Machine Learning: A Probabilistic Perspective", the MIT Press Cambridge, Massachusetts London, England.

[3] Marc Peter Deisenroth, A Aldo Faisal, Cheng soon Ong, "Mathematics for Machine Learning", Cambridge University Press.

[4] Li M. Chen, Zhixun Su, Bo Jiang, "Mathematical problems in data science, Theoretical and Practical Methods", Springer (2015).