AIML - CS 337

Lecture 3: Overview of Linear Algebra for ML

August $8^{th}, 2022$

Lecturer: Abir De *Scribe: Prathamesh, Utkarsh Ranjan, Harshvardhan Ahirwar, Ashutosh Pankaj*

Linear Algebra is one of the key fundamental concepts required in Machine Learning. Few of the key uses of Linear Algebra is in its use of modelling a system as a linear transformation of input data, helping solve a system of linear equations, and in also representing data with smaller number of parameters , using Principal Component Analysis, **PCA**.

1 Vectors and their Properties

Vectors are the basic representation unit of a data-point , an ordered tuple of numbers . In ML (and even in generally) we will deal with vectors whose entries lie in We start with few basic properties, functions and notations for a vector ($\in \mathbb{R}^n$).

1.1 Dot product

Definition 1.1. The dot product of two equidimensional vectors **u , v** $\in \mathbb{R}^n$ is given by

$$\begin{aligned}\mathbf{u} \cdot \mathbf{v} &= \sum u_i v_i \\ &= \mathbf{u^T v}\end{aligned}$$

where x_i denotes the i^{th} component of vector x, and x^T represent the transpose of a vector. In the above definition of dot product, it has been assumed that the vectors are represented as columns.

1.2 Independence

A set of vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ is said to be linearly independent if one cannot be represented by any linear combination of other vectors. More formally,

Definition 1.2. a set of vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ is said to be linearly independent iff for scalars c_i's we have,

$$\sum c_i \mathbf{v_i} = 0 \rightarrow c_i = 0 \qquad \forall i \in [1, n]$$

The above definition can also be seen that the solution of the equation

$$\mathbf{Ac} = \mathbf{0}$$

is only the trivial solution, where is the matrix $\mathbf{A} = [\mathbf{v_1} \; \mathbf{v_2} \; \ldots \; \mathbf{v_n}]$

1.3 Vector space

A set of vectors **V** is said to qualify as a vector space if it is closed under the operations of addition and scalar multiplication i.e. :

Definition 1.3. A set of vectors $\mathbf{V}$, is said to be a vector space , if for any two vectors $\mathbf{u}, \mathbf{v} \in \mathbf{V}$ we have

$$a * \mathbf{u} \in \mathbf{V}, \qquad a \in \mathbb{R} \tag{1}$$

$$a * \mathbf{u} + b * \mathbf{v} \in \mathbf{V}, \qquad a, b \in \mathbb{R} \tag{2}$$

A subset of independent vectors of a vector space of highest cardinality is called a **basis** for the vector space. Every vector in the vector space can be represented as a linear combination of some or all vectors in a basis of the vector subspace. Note that the basis need not be unique for the vector space.

2 Matrices

A matrix is a rectangular array or a table of numbers, symbols , or expressions, arranged in rows and columns, which is used to represent a mathematical object or a property of such an object. Matrices forms a fundamental part of linear algebra which have geometric meaning asscocaited to it while dealing with data in Machine Learning.
An N $\times$ M matrix has N rows and M columns. Following is an example of a 2x2 matrix A.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

2.1 Matrix Multiplication

Definition 2.1. If V is a matrix of size n $\times$ m and W is an m $m \times p$ matrix, then the matrix product U = VW is a size of $n \times p$ and can be computed as follows:

$$V = \begin{bmatrix} v_{1,1} & \dots & v_{1,m} \\ \vdots & \vdots & \vdots \\ v_{n,1} & \dots & v_{n,m} \end{bmatrix}, W = \begin{bmatrix} w_{1,1} & \dots & w_{1,p} \\ \vdots & \vdots & \vdots \\ w_{m,1} & \dots & w_{m,p} \end{bmatrix}, U = \begin{bmatrix} u_{1,1} & \dots & u_{1,p} \\ \vdots & \vdots & \vdots \\ u_{n,1} & \dots & u_{n,p} \end{bmatrix}$$

where $u_{i,j} = \sum_{k=1}^{m} v_{i,k} w_{k,j}$

2.2 Determinants

Determinant is a scalar value that is a function of the entries of a square matrix. In the case of a 2x2 matrix the determinant can be defined as

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

2.3 Linear Systems

A linear system or a system of linear equation is a collection of one or more linear equations involving the same variables. A general system of linear equation with n unknows and coefficients can be written as

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned}
$$

We can represent a system of linear equations using a coefficient matrix A, a vector of variables $\mathbf{x}$ and constants $\mathbf{b}$ as : $\mathbf{Ax} = \mathbf{b}$

3 Solving System of Equations

For a system of linear equations in n variables can have the following solutions :

- Exactly one solution
- No solution
- An infinite number of solutions

3.1 Matrix Inversion

Given a linear system of equation represented by Ax = b, we can find a solution x = $A^{-1}b$, where A^{-1} is called the inverse of the matrix.

- If $\mathbf{A}$ is invertible we can find unique solution to our systems of equation.
- **Gauss-Jordan :** We perform elimination step on the augmented matrix [A I] to give the augemnted matrix [I A^{-1}]. Gauss-Jordan elimination gives a Reduced Row Echelon Form.

3.2 Gaussian Elimination

Given a system of linear equations denoted by

$$\mathbf{A}x = \mathbf{b}$$

we can solve the equations using Gauss Jordan Elimination. We note the elementary row operations are given as swapping two rows in the matrix , multiplying rows by a constant factor , and adding two rows. After performing Gaussian Elimination on the coefficient matrix we get a Row Echelon Form.

Algorithm 1: Gaussian Elimination

```
Data: A, b
Result: x such that Ax = b
A_b ← A|b;
n ← A.shape[0];
for i in 1 to n-1 do
    if A_b[i][i] is 0 then
        X ← "No solution exists";
    else
        for j in i+1 to n do
            for k in 1 to n+1 do
                A_b[j][k] ← A_b[j][k] − A_b[j][i]/A_b[i][i] ∗ A_b[i][k];
            end
        end
    end
end
X_n ← A_b[n][n + 1]/A_b[n][n];
for i in n-1 to 1 do
    X_i ← A_b[i][n + 1];
    for j in i+1 to n do
        X_i ← X_i − A_b[i][j] ∗ X_j;
    end
    X_i ← X_i/A_b[i][i];
end
```

3.3 LU decomposition and application

We can also solve for the system of linear equations $Ax = b$, by first converting $\mathbf{A}$ into a product of two matrices, such that one is completely a lower triangular matrix L and other is an upper triangular matrix U,

$$A = LU$$

Matrix L and U , can be found by the help of Gaussian Elimination applied on the matrix A. Once we have L and U , we can then solve for x by first solving the equation ,

$$Ly = b$$

for y , and then the equation

$$Ux = y$$

4 Column space, Null space and Invertibility

4.1 Column space :

Column space of a matrix A, or $C(A)$, is the space consisting of all possible linear combinations of the columns of A, and for any real vector x, Ax will lie in the column space of A.

- If there exists a solution for the equation $b = Ax$ for a real matrix A and real vectors b and x, then b lies in the $C(A)$ and vice versa.

4.2 Null space :

Null space of a matrix A, or $N(A)$, is the space spanned by all the solutions x, of $Ax = 0$.

The solutions form a vector space because :

- If $Ax = 0$ then $A(cx) = c(Ax) = 0$
- If $Ax = 0$ and $Ay = 0$ then $A(x + y) = 0$

Having a vector x other than the null vector, such that $Ax = 0$, implies that the columns of matrix A are dependent.

4.3 Rank of a matrix :

Rank of a matrix A is defined as the size of the maximal set of independent columns of the matrix A, and those columns form a basis for $C(A)$.

- If A^{-1} exists, the only solution to $Ax = b$ is $x = A^{-1}b$
- In addition to the first statement, A is singular iff there are solutions other than $x = 0$ to $Ax = 0$, or in other words, iff it has a non-singular null-space $N(A)$.

A matrix A is called a full column rank matrix if all the columns in A are independent.

4.4 Invertibility :

A square matrix A is invertible if there exists a square matrix B such that $AB = BA = I_n$, and B is known as inverse of A, also denoted by A^{-1}.

- A square matrix is invertible iff it is a full column rank matrix.

4.5 Computing the Inverse - Gauss Jordan Elimination :

The Gauss-Jordan elimination method addresses the problem of solving several linear systems $Ax_i = b_i (1 \leq i \leq N)$ at once, such that each linear system has the same coefficient matrix A but a different right hand side b_i.

We have seen how elimination matrices are used to convert a coefficient matrix A into some upper triangular matrix U,

$$U = E_{32}(E_{31}(E_{21}A)) = (E_{32}E_{31}E_{21})A$$

Now, further apply elimination steps until U gets transformed into identity matrix:

$$I = E_{13}(E_{12}(E_{23}(E_{32}(E_{31}(E_{21}A))))) = (E_{13}E_{12}E_{23}E_{32}E_{31}E_{21})A$$

By definition, $X = (E_{13}E_{12}E_{23}E_{32}E_{31}E_{21})$ must be A^{-1}

Note that the above method works only if A is invertible.

So, Gauss-Jordan is basically elimination steps over the augmented matrix $[AI]$ (representing the equation $AX = I$) to give the augmented matrix $[IA^{-1}]$ (representing the equation $IX = A^{-1}$)

4.6 Dealing with Rectangular matrices :

What if A is not a square matrix but rather a rectangular matrix of size mxn, such that $m \neq n$. Does there exist a notion of A^{-1}? The answer depends on the rank of A.

- If A is full row rank and $n > m$, then AA^T is a full rank mxm matrix $\iff (AA^T)^{-1}$ exists with $A^T(AA^{T-1})$ leading to I and is therefore called the right inverse of A. When the right inverse of A is multiplied on its left, we get the projection matrix $A^T(AA^T)^{-1}A$, which projects matrices onto the row space of A.
- If A is full column rank and $m > n$, then A^TA is full rank nxn matrix $\iff (A^TA)^{-1}$ exists with $(A^TA)^{-1}A^T$ leading to I and is therefore called the left inverse of A. When the left inverse of A is multiplied on its right, we get the projection matrix $A(A^TA)^{-1}A^T$, which projects matrices onto the column space of A.

If A is a full column rank matrix (that is, its columns are independent), A^TA is invertible.

We will show that the null space of A^TA is 0, which implies that the square matrix A^TA is full column (as as well as row) rank is invertible. That is if $A^TAx = 0$, then $x = 0$. Note that if $A^TAx = 0$, then $x^TA^TAx = ||Ax|| = 0$ which implies that $Ax = 0$. Since the columns of A are linearly independent, its null space is 0 and therefore, $x = 0$.