A Brief Intro to Linear Algebra

Ishwar Sethi
CSI 5810

Why Linear Algebra?

Linear algebra provides a convenient mathematical representation for data and operations on data

The word algebra stems from the Arabic word *al-jabr*, which has its roots in the title of a 9th century manuscript written by the mathematician Al-Khwarizmi.

Baudhayana, author of the Baudhayana Sulba Sutra, a Vedic Sanskrit geometric text, contains quadratic equations, and calculates the square root of 2 correct to five decimal places, 800 BC [From Wikipedia]

Let's Begin with Scalers

- A scaler is a number
- Examples of scalers in data mining:
  - Age of an employee
  - Education in years
  - Salary

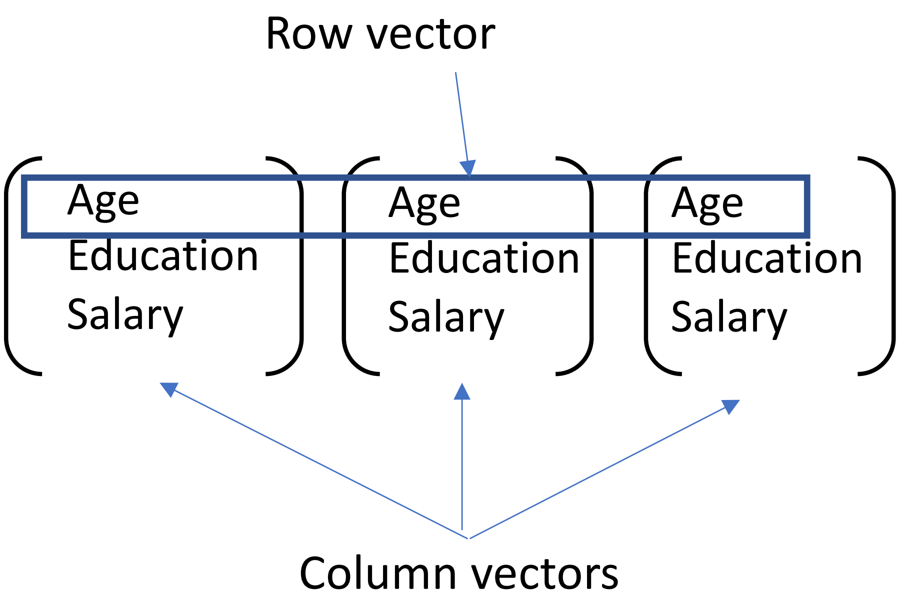
Vectors

- A vector is a collection of scalers organized vertically or horizontally

$$\begin{pmatrix} \text{Age of an employee} \\ \text{Education in years} \\ \text{Salary} \end{pmatrix}$$

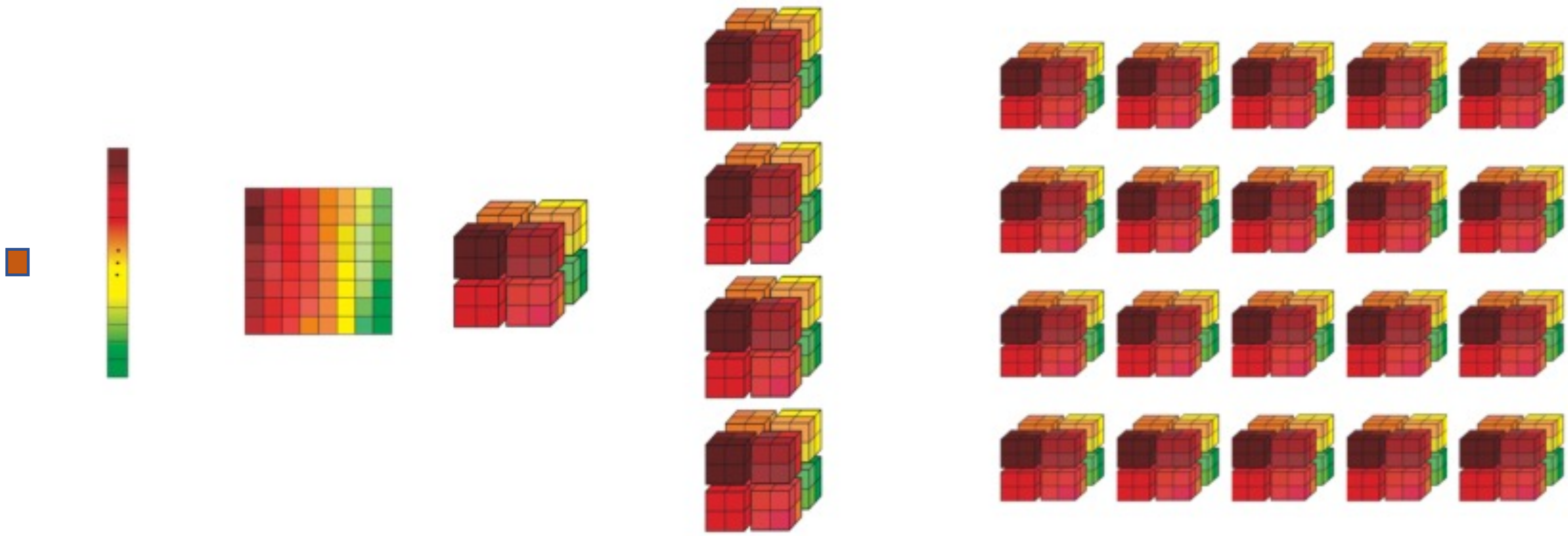
Matrices

- A matrix is a collection of vectors



Tensors

- A tensor is a collection of matrices.



Scalar -> Vector -> Matrix -> Tensor

Matrix Representation

- A matrix is represented as a rectangular array of numbers (or functions).

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ a_{m1} & & & a_{mn} \end{bmatrix}$$

- The matrix shown above is of size $m{\times}n$, meaning the matrix has m rows and n columns.
- The elements of a matrix, here represented by the letter ‘a’ with subscripts, can consist of numbers, variables, or functions of variables.

Vectors

- A vector is simply a matrix with either one row or one column. A matrix with one row is called a *row vector*, and a matrix with one column is called a *column vector*.
- Transpose: A row vector can be changed into a column vector and vice-versa by taking the *transpose* of that vector. e.g.

$$\text{if} \quad A = \begin{bmatrix} 3 & 4 & 5 \end{bmatrix} \quad \text{then} \quad A^T = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

Matrix Addition

- Matrix addition is only possible between two matrices which have the same size.
- The operation is done simply by adding the corresponding elements. e.g.:

$$\begin{bmatrix} 1 & 3 \\ 4 & 7 \end{bmatrix} + \begin{bmatrix} 6 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 5 \\ 7 & 8 \end{bmatrix}$$

Matrix Scalar Multiplication

- Multiplication of a matrix or a vector by a scalar is also straightforward:

$$5 * \begin{bmatrix} 1 & 3 \\ 4 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 20 & 35 \end{bmatrix}$$

Transpose of a matrix

- Taking the transpose of a matrix is like that of a vector:

$$\text{if} \quad A = \begin{bmatrix} 1 & 3 & 8 \\ 4 & 7 & 2 \\ 6 & 5 & 0 \end{bmatrix}, \quad \text{then} \quad A^T = \begin{bmatrix} 1 & 4 & 6 \\ 3 & 7 & 5 \\ 8 & 2 & 0 \end{bmatrix}$$

- The diagonal elements in the matrix are unaffected, but the other elements are switched. A matrix which is the same as its own transpose is called *symmetric*, and one which is the negative of its own transpose is called *skew-symmetric*.

- Symmetric: $\mathbf{A}^T = \mathbf{A}$.
- Skew-symmetric: $\mathbf{A}^T = -\mathbf{A}$.
- Examples:

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 5 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & 3 \\ 2 & -3 & 0 \end{bmatrix}$$

symmetric   skew-symmetric

Matrix Multiplication

- The multiplication of a matrix into another matrix is not possible for all matrices, and the operation is ***not commutative***:

$$AB \neq BA \text{ in general}$$

- In order to multiply two matrices, the first matrix must have the same number of columns as the second matrix has rows.
- So, if one wants to solve for C=AB, then the matrix A must have as many columns as the matrix B has rows.
- The resulting matrix C will have the same number of rows as did A and the same number of columns as did B.

Inner Product

The *inner product* of two vectors

$$\mathbf{a}' = [\mathrm{a}_1 \quad \mathrm{a}_2 \quad \cdots \quad \mathrm{a}_n]$$

and $\mathbf{x} = \begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \\ \vdots \\ \mathrm{x}_n \end{bmatrix}$ is defined as

$$\mathbf{a}'\mathbf{x} = \mathrm{a}_1\mathrm{x}_1 + \mathrm{a}_2\mathrm{x}_2 + \cdots + \mathrm{a}_n\mathrm{x}_n = \sum_{i=1}^{n} \mathrm{a}_i\mathrm{x}_i$$

‘ stands for transpose

Also known as *dot product*

Matrix Multiplication

- The operation is done as follows:

  using index notation:

$$C_{jk} = \sum_{l=1}^{n} A_{jl} B_{lk}$$

- for example:

$$AB = \begin{bmatrix} 4 & 3 \\ 7 & 2 \\ 9 & 0 \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 4*2+3*1 & 4*5+3*6 \\ 7*2+2*1 & 7*5+2*6 \\ 9*2+0*1 & 9*5+0*6 \end{bmatrix}$$

$$= \begin{bmatrix} 11 & 38 \\ 16 & 47 \\ 18 & 45 \end{bmatrix}$$

Linear systems of equations

- One of the most important application of matrices is for solving linear systems of equations which appear in many different problems including electrical networks, statistics, and numerical methods for differential equations.
- A linear system of equations can be written:

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

- This is a system of *m* equations and *n* unknowns.

Linear systems cont’

- The system of equations shown on the previous slide can be written more compactly as a matrix equation:

  Ax=b

- where the matrix A contains all the coefficients of the unknown variables from the LHS, x is the vector of unknowns, and b a vector containing the numbers from the RHS

Determinants

- Determinants are useful in eigenvalue problems and differential equations.
- Can be found only for square matrices.
- Simple example: 2nd order determinant

$$\det A = \begin{vmatrix} 1 & 3 \\ 4 & 7 \end{vmatrix} = 1*7 - 3*4 = -5$$

3^{rd} order determinant

- The determinant of a 3X3 matrix is found as follows:

$$\det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

- The terms on the RHS can be evaluated as shown for a 2^{nd} order determinant.

Matrix rank

- The rank of a matrix is simply the number of independent row vectors in that matrix.
- The transpose of a matrix has the same rank as the original matrix.

$$\begin{pmatrix} 3 & 2 & 4 \\ -1 & 1 & 2 \\ 9 & 5 & 10 \end{pmatrix}$$

Rank ?

Matrix inverse

- The inverse of the matrix A is denoted as A^{-1}
- By definition, $AA^{-1} = A^{-1}A = I$, where I is the identity matrix.
- Theorem: The inverse of an nxn matrix A exists if and only if the rank A = n.
- Inverse exists only if the determinant of the matrix is nonzero.
- $(AB)^{-1} = B^{-1}A^{-1}$

Singular Matrix

- Let the square matrix $\mathbf{A}_{n\times n}$ be of order n.
- The matrix $\mathbf{A}$ is said to be singular if any of the following equivalent conditions exists:

i. $|\mathbf{A}| = 0$

ii. If a particular row (column) can be formed as a linear combination of the other rows (columns)

iii. $r(\mathbf{A}) < n$

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ -6 \\ 9 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 5 \\ -5 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{x}_4 = \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix} \quad \mathbf{x}_5 = \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}$$

$$2\mathbf{x}_1 + 3\mathbf{x}_2 - 3\mathbf{x}_3 = \begin{bmatrix} 6 \\ -12 \\ 18 \end{bmatrix} + \begin{bmatrix} 0 \\ 15 \\ -15 \end{bmatrix} + \begin{bmatrix} -6 \\ -3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}$$

$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$, are linearly dependent vectors.

Linearly Independent Vectors

If there exists a vector $\mathbf{a} \neq \mathbf{0}$, such that

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{0}$$

then provided none of the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is null, those vectors (the columns of **X**) are said to be *linearly* *ıdependent vectors.*

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ -6 \\ 9 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 5 \\ -5 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{x}_4 = \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix} \quad \mathbf{x}_5 = \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}$$

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = \begin{bmatrix} 3a_1 \\ -6a_1 \\ 9a_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 5a_2 \\ -5a_2 \end{bmatrix} = \begin{bmatrix} 3a_1 \\ -6a_1 + 5a_2 \\ 9a_1 - 5a_2 \end{bmatrix}$$

There are no values a_1 and a_2, which makes it a null vector other than $a_1 = a_2 = \mathbf{0}$. Therefore, $\mathbf{x}_1$ and $\mathbf{x}_2$ are *linearly independent* vectors.

Trace of a Matrix

- The sum of the diagonal elements of a square matrix is called the *trace* of the matrix, that is,

$$\mathrm{tr}(\mathbf{A}) = a_{11} + a_{22} + \ldots + a_{nn}$$

- Example of trace

Let $\mathbf{A} = \begin{bmatrix} 1 & 5 & 9 \\ -3 & 2 & 8 \\ 4 & 7 & 6 \end{bmatrix}$

Then $\mathrm{tr}(\mathbf{A}) = 1 + 2 + 6 = 9$

Identity Matrix

- A diagonal matrix having all diagonal elements equal to one is called an *identity* matrix.
- It is denoted by the letter $\mathbf{I}$.
- For any matrix $\mathbf{A}$ of order $m \times n$,

$$\mathbf{I}_m\mathbf{A}_{m\times n} = \mathbf{A}_{m\times n}\mathbf{I}_n = \mathbf{A}_{m\times n}$$

Example of Identity Matrix

$$\mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is an identity matrix of order 4.

Special matrices

- A matrix is called *symmetric* if:

$$A^T = A$$

- A skew-symmetric matrix is one for which:

$$A^T = -A$$

- An *orthogonal* matrix is one whose transpose is also its inverse:

$$A^T = A^{-1}$$

Is $A = \begin{bmatrix} -0.55978 & -0.80992 & 0.17514 \\ 0.55713 & -0.52432 & -0.64396 \\ 0.61339 & -0.2629 & 0.74474 \end{bmatrix}$ orthogonal?

Eigenvalues and Eigenvectors

- Let A be an *nxn* matrix and consider the vector equation:

$$Ax = \lambda x$$

- A value of λ for which this equation has a solution $x \neq 0$ is called an eigenvalue of the matrix A.
- The corresponding solutions x are called the eigenvectors of the matrix A.
- Conceptually, the eigenvectors of a matrix point to the directions of some underlying property captured in the matrix and eigenvalues indicate the level/strength of the property.

Solving for eigenvalues

$$Ax=\lambda x$$

$$Ax - \lambda x = 0$$

$$(A- \lambda I)x = 0$$

- This is a *homogeneous* linear system, homogeneous meaning that the RHS are all zeros.
- For such a system, a theorem states that a solution exists given that det(A- λI)=0.
- The eigenvalues are found by solving the above equation.

Solving for eigenvalues cont'

- Simple example: find the eigenvalues for the matrix:

$$A = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}$$

- Eigenvalues are given by the equation det(A-λI) = 0:

$$\det(A - \lambda I) = \begin{vmatrix} -5-\lambda & 2 \\ 2 & -2-\lambda \end{vmatrix}$$

$$= (-5-\lambda)(-2-\lambda) - 4 = \lambda^2 + 7\lambda + 6$$

- So, the roots of the last equation are -1 and -6. These are the *eigenvalues* of matrix A.

Eigenvectors

- For each eigenvalue, λ, there is a corresponding eigenvector, x.
- This vector can be found by substituting one of the eigenvalues back into the original equation: $Ax = \lambda x$: for the example:

$$-5x_1 + 2x_2 = \lambda x_1$$

$$2x_1 - 2x_2 = \lambda x_2$$

- Using λ=-1, we get $x_2 = 2x_1$, and by arbitrarily choosing $x_1 = 1$, the eigenvector corresponding to λ=-1 is:

$$v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

and similarly,

$$v_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Properties of Eigenvalues & Eigenvectors

For $\mathbf{Ax} = \lambda\mathbf{x}$,

1. $\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}$ and $\mathbf{A}^{-1}\mathbf{x} = \lambda^{-1}\mathbf{x}$, when $\mathbf{A}$ is nonsingular.
2. $c\mathbf{Ax} = c\lambda\mathbf{x}$ for any scalar c.
3. $f(\mathbf{A})\mathbf{x} = f(\lambda)\mathbf{x}$ for any polynomial function $f(\mathbf{A})$

For $\mathbf{Ax} = \lambda\mathbf{x}$,

4. $\sum_{i=1}^{n} \lambda_i = \mathrm{tr}(\mathbf{A})$ $\qquad$ $\prod_{i=1}^{n} \lambda_i = |\mathbf{A}|$

the sum of eigenvalues of a matrix equals its trace,

and their product equals its determinant.

For $\mathbf{Ax} = \lambda\mathbf{x}$,

5. If $\mathbf{A}$ is symmetric, then
   - the eigenvalues of matrix $\mathbf{A}$ are all real
   - $\mathbf{A}$ is diagonable
   - the eigenvectors are orthogonal to each other
   - the rank of $\mathbf{A}$ equals the number of nonzero eigenvalues
   - positive definite matrices have eigenvalues all greater than zero and vice versa

Quadratic Form

- A *quadratic form* is the product of a row vector $\mathbf{x}'$, a matrix $\mathbf{A}$, and the column vector $\mathbf{x}$, that is, $\mathbf{x}'\mathbf{A}\mathbf{x}$.
- This is a quadratic function of the x's. Notice that to result in the same quadratic function of x's, you can use many different matrices.
- Each matrix has the same diagonal elements, and the sum of each pair of symmetrically placed off-diagonal elements a_{ij} and a_{ji} is the same.

Example of a Quadratic Form

When $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 3 & 1 & 4 \end{bmatrix}$

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 3 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= x_1^2 + 5x_2^2 + x_3^2 + 4x_1x_2 + 6x_2x_3 + 2x_2x_3$$

Positive Definite Quadratic Form and Positive Definite Matrix

- When $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x}$ other than $\mathbf{x} = \mathbf{0}$, then $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a *positive definite* quadratic form;
- $\mathbf{A} = \mathbf{A}'$ is correspondingly a *positive definite matrix*.

Example of Positive Definite Quadratic Form and Matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

The matrix A is positive definite

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= 2x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_2x_3 + 2x_2x_3$$

$$= (x_1 + 2x_2)^2 + (x_1 + x_3)^2 + (x_2 + x_3)^2 > 0$$

other than $\mathbf{x} = 0$

Positive Semidefinite Quadratic Form and Positive Semidefinite Matrix

- When $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x}$ and $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$ for some $\mathbf{x} \neq \mathbf{0}$, then $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a *positive semidefinite* quadratic form;
- $\mathbf{A} = \mathbf{A}'$ is correspondingly a *positive semidefinite matrix*.

Let $\mathbf{A} = \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix}$

Then $\mathbf{x}'\mathbf{A}\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$= 37x_1^2 + 13x_2^2 + 17x_3^2 - 4x_1x_2 - 48x_2x_3 - 6x_2x_3$$

$$= (x_1 - 2x_2)^2 + (6x_1 - 4x_3)^2 + (3x_2 - x_3)^2$$

This is zero for $\mathbf{x}' = [2 \;\; 1 \;\; 3]$ and for any scalar multiple thereof, as well as for $\mathbf{x} = \mathbf{0}$.

For in depth study of linear algebra

https://math.mit.edu/~gs/linearalgebra/

For a basic intro

https://towardsdatascience.com/linear-algebra-for-deep-learning-f21d7e7d7f23