CSO507
Deep Learning

Overview of Linear Algebra

Koustav Rudra

11/01/2023

Scalars

- A scalar is a single number
- It can be real, integer, etc.
- Typically it will be denoted using lowercase italics: a, x, n
- Example:
  - Let $s \in \mathbb{R}$ be the slope of the line
  - Let $n \in \mathbb{N}$ be the number of units

Vectors

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

- It is an array of numbers (e.g., scalars) and arranged in order
- Typically it will be denoted using lowercase bold font: $\boldsymbol{x}$
- Need to specify what kind of numbers are stored
  - If each element is in $\mathbb{R}$ then the vector lies in $\mathbb{R}^n$ (Cartesian product n times)
- Identify points in space, each element giving the coordinate along different axis
- A set of elements, x_1, x_2, x_3 can be specified as x_S where $S = 1,3,5$
- x_{-2} is a vector containing all elements except x_2

Matrices

- A matrix is a 2D array of numbers: $\boldsymbol{X} = [x_{i,j}] = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix}$

- Example notation for type and shape: $\boldsymbol{X} \in \mathbb{R}^{m \times n}$

- The jth column will be denoted as: $\boldsymbol{x}_j$ or $X_{:,j}$ $\boldsymbol{X} = \begin{bmatrix} | & | & \cdots & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \\ | & | & \cdots & | \end{bmatrix}$

- n-dimensional vector can be represented as n rows and 1 column: $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

Tensors

- A tensor is an array of numbers that may have
  - Zero dimensions, and be a scalar

```
>>> torch.tensor(3.14159)
tensor(3.1416)
```

  - One dimension, and be a vector

```
>>> torch.tensor([0, 1])
tensor([ 0,  1])
```

  - Two dimensions, and be a matrix

```
>>> torch.tensor([[0.1, 1.2], [2.2, 3.1], [4.9, 5.2]])
tensor([[ 0.1000,  1.2000],
        [ 2.2000,  3.1000],
        [ 4.9000,  5.2000]])
```

  - Or, more dimensions

Matrix Transpose

- Rows and columns are interchanged that is $\boldsymbol{X}^T = [\boldsymbol{X}_{i,j}]^T = [\boldsymbol{X}_{j,i}]$

- For example, $\boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \end{bmatrix} \quad \boldsymbol{X}^T = \begin{bmatrix} x_{1,1} & x_{2,1} \\ x_{1,2} & x_{2,2} \\ x_{1,3} & x_{2,3} \end{bmatrix}$

- Mirror image of matrix across the main diagonal

- For scalars, $a = a^T$

- $(\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T$

Matrix Manipulation

- Matrix addition: $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} \Longrightarrow\ C_{i,j} =\ A_{i,j} +\ B_{i,j}$

- Matrix multiplication: $\boldsymbol{C} = \boldsymbol{A} \times \boldsymbol{B} \Longrightarrow C_{i,j} = \sum_k A_{i,k} \times B_{k,j}$

- Multiplication and addition are associative:
  - **(AB)C = A(BC)**
  - **A + (B + C) = (A + B) + C**

- Multiplication is distributive
  - $\boldsymbol{A} \times (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$

- Multiplication is not commutative
  - $\boldsymbol{AB} \neq \boldsymbol{BA}$

Matrix Product

- Let us assume $Z = X \times Y$, where $X \in \mathbb{R}^{m \times n}$ and $Y \in \mathbb{R}^{n \times p}$
- Number of columns in X should be equal to number of rows in Y
- $Z_{i,j} = \sum_{k=1}^{n} X_{i,k} \times Y_{k,j}$

Identity Matrix

- All elements are 0 except for diagonal elements which are 1

- Example: $\boldsymbol{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

- $\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x}$

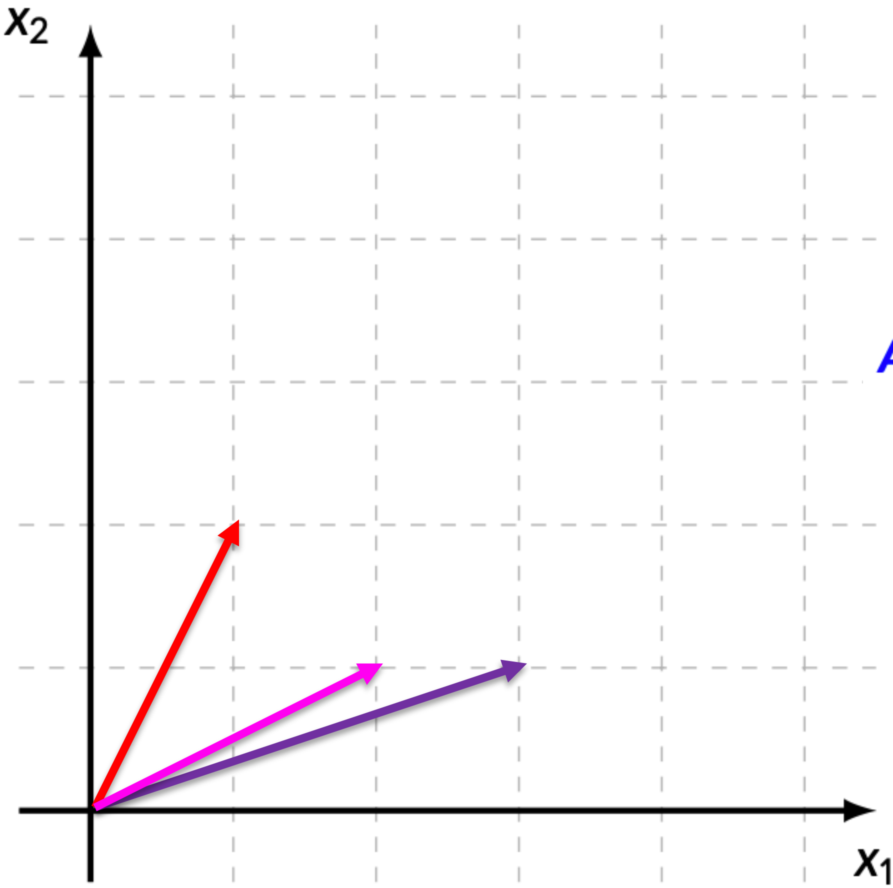
Systems of Equations

- Consider following equations:
  - $4x_1 - 5x_2 = -13$
  - $-2x_1 + 3x_2 = 9$

- Can be expressed in the form $\boldsymbol{Ax} = \boldsymbol{b}$
  - $\boldsymbol{A} = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}, \boldsymbol{b} = \begin{pmatrix} \mathbf{-13} \\ \mathbf{9} \end{pmatrix}$

- A linear system of equations may have –
  - No solution
  - Many solutions
  - Exactly one solution
    - This means multiplication by the matrix is an invertible function

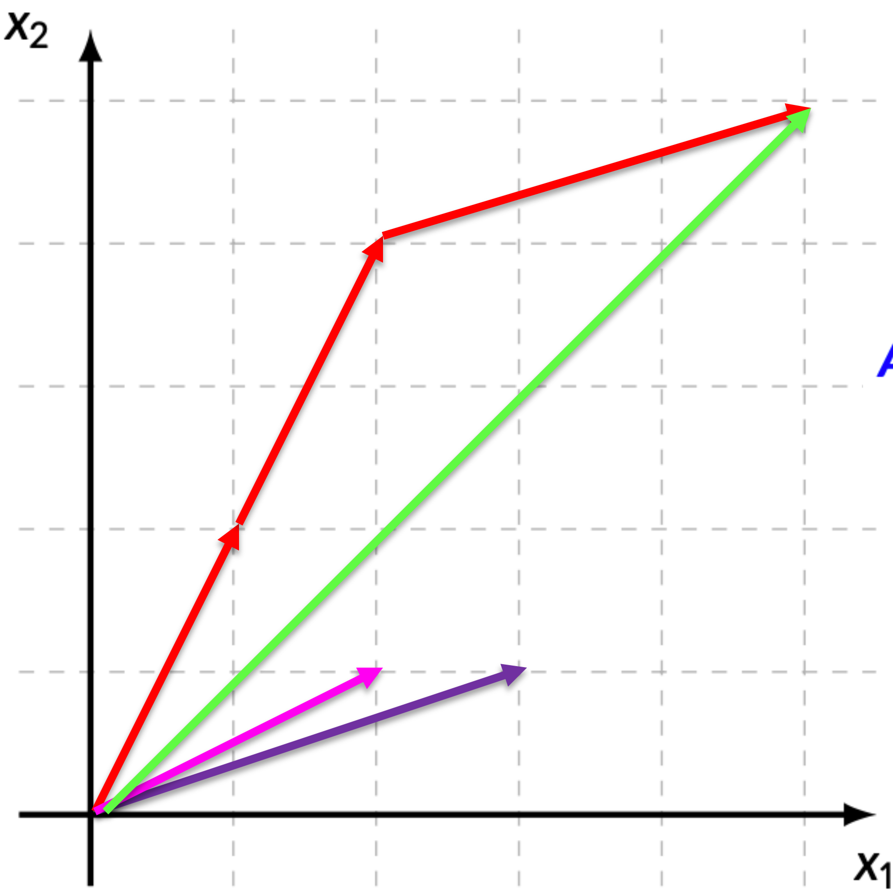
Linear Transformation



$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\boldsymbol{Ax} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times 2 + \begin{bmatrix} 3 \\ 1 \end{bmatrix} \times 1 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

Linear Transformation



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

$$Ax = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times 2 + \begin{bmatrix} 3 \\ 1 \end{bmatrix} \times 1 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

Matrix Inversion

- $\boldsymbol{A^{-1} \times A = I_n}$
- Solving a system of equations using inverse
  - $\boldsymbol{Ax = b}$
  - $\boldsymbol{A^{-1}Ax = A^{-1}b}$
  - $\boldsymbol{I_n x = A^{-1}b}$
  - $\boldsymbol{x = A^{-1}b}$

- Numerically unstable, but useful for abstract analysis
- Matrix cannot be inverted if
  - More rows than columns
  - More columns than rows
  - Redundant rows/columns
    - (“linearly dependent", "low rank”)

Linear Independence

- Column can be thought of as specifying direction from origin

- Each element of $\boldsymbol{x}$ specify how far we should move in each of these direction i.e., $\boldsymbol{Ax} = \sum x_i A_{:,i}$

- Formally, this is a linear combination of the set of vectors

- Span of set of vectors is the set of all points obtainable by linear combination of the original vectors

- What does $\boldsymbol{Ax} = \boldsymbol{b}$ represent?
  - Testing whether b is in span of column of A
  - Span is known as column space or range of A

Linear Independence

- A set of vectors is linearly independent if no vectors in the set is a linear combination of of other vectors
  - No new points will be added if linear combination of vectors are added in the set
- Suppose column space is $\mathbb{R}^m$
  - Need to have exactly *m* linearly independent column
  - No set of *m* dimensional vectors can have more than *m* mutually linearly independent column

- A square matrix with linearly dependent columns is known as singular

- A matrix to have inverse
  - $Ax = b$ has at most one solution for each value of b

- *Is it possible to solve $Ax = b$ if A* is not square but singular
  - Yes but matrix inversion method cannot be used

Norms

- Measure the size of vector. It is defined as

$$L^p = \|\mathbf{x}\|_p = \left(\sum_i |x_i|^p\right)^{1/p} \quad p \in \mathbb{R}, p \geq 1$$

- Intuitive meaning: distance of x from the origin

- Norm is any function f that satisfies
  - $f(x) = 0 \implies x = 0$
  - $f(x + y) \leq f(x) + f(y) \implies Triangle\ Inequality$
  - $\forall \alpha \in \mathbb{R}, f(\alpha x) = \alpha f(x)$

Norms

- L^2 norm is known as Euclidean norm
  - It is often denoted as $\| x \|$ instead of $\|x\|_2$
  - Squared L^2 norm can be determined by $x^T x$

- Derivative of the squared L^2 norm depend only on the corresponding element

$$\frac{\partial(\mathbf{x}^T\mathbf{x})}{\partial x_i} = \frac{\partial(x_1^2 + \ldots + x_m^2)}{\partial x_i}$$

- Derivative of L^2 depend on entire vector

$$\frac{\partial\sqrt{(\mathbf{x}^T\mathbf{x})}}{\partial x_i} = \frac{\partial\sqrt{(x_1^2 + \ldots + x_m^2)}}{\partial x_i}$$

- Square L^2 norm is undesirable as it increases very slowly at the origin

Norms

- Need to identify elements that are zero and elements that are non-zero but small
  - Need a function that grow at the same rate in all locations
  - L^1 can be used to differentiate zero and non-zero elements

$$L^1 = \|\boldsymbol{x}\| = \sum_i |x_i|$$

- L^∞ (max norm) - Absolute value of the elements with the largest magnitude in the vector

$$\|\boldsymbol{x}\|_\infty = \max_i |x_i|$$

- Frobenius norm

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$$

  - This is analogous to L^2 norm of vector

Special Matrices

- Diagonal matrices — Non-zero diagonal elements and rests are zero
  - $D_{i,j} = 0, i \neq j$
  - Identity matrix
  - diag(v) — vectors using diagonal elements
  - diag(v)x — x_i is scaled by v_i
  - Inversion is easy **$\text{diag}(v)^{-1} = \text{diag}([1/v_1, 1/v_2, \ldots, 1/v_n]^T)$**

- Are the diagonal matrices need to be square?
  - No
  - Dx — Scaling each element of x
    - Concatenate some zero if D is taller
    - Discard some last elements if D is wider

- Symmetric matrix — Arises when the entries are generated by a function of two arguments that does not depend on order

Special Vectors & Matrices

- Unit vector — A vector with unit norm $\|x\|_2 = 1$

- For vectors x and y, if $x^T y = 0$
  - Norm of x or y is zero
  - x and y are at 90°

- In $\mathbb{R}^n$, at most n vectors may be mutually orthogonal with non-zero norm

- Vectors orthogonal and have unit norm is known as orthonormal

- Orthogonal matrix — Square matrix, rows are mutually orthonormal, columns are mutually orthonormal
  - $A^T A = A A^T = I \Longrightarrow A^T = A^{-1}$
  - Orthonormal matrices are of interest as inverse computation is easy