Lecture 2

Math Foundations Team

BITS Pilani

Pilani | Dubai | Goa | Hyderabad

Vector Spaces

- We would like to shift focus from systems of linear equations and their solutions to vector spaces.
- Vector spaces are structured spaces in which vectors live.
- What do we mean by "structure" here?

Groups

- We already have some notion about the structure of a vector space, i.e adding two vectors returns a vector, multiplying a vector by a scalar also returns a vector.
- To formalize these notions, we need the concept of a *Group*
- *A Group is a set G and an operation $\otimes : G \times G \to G$ defined on G. Then $(G, \otimes)$ is called a group if the following properties hold:*
  - *Closure of G under $\otimes$:* $\forall x, y \in G, x \otimes y \in G$
  - *Associativity:* $\forall x, y, z \in G, (x \otimes y) \otimes z = x \otimes (y \otimes z)$
  - *Neutral or Identity element:* $\ni e \in G, \forall x \in G x \otimes e = x$
  - *Inverse element:* $\forall x \in G, \ni y \in G, x \otimes y = y \otimes x = e$
- *If in addition to all the above properties we have* $\forall x, y \in G, x \otimes y = y \otimes x$ *then* $(G, \otimes)$ *is an Abelian group*

Some examples

- $(Z, +)$ is an Abelian group. We can see that all the aforementioned properties hold.
- What about $(N_0, +)$? This is not a group since there is no inverse element for an arbitrary element in it.
- $(Z, .)$ where Z is the set of integers, and . is product. It has the identity element but there exist elements in it that don't have an inverse.

Vector Spaces

- We will now consider sets that are just like groups in terms of its properties with respect to an inner operation $+$, and also have an outer operation called . which denotes the multiplication of a vector $x \in G$ by a scalar $\lambda \in R$.
- The inner operation can be viewed as a form of addition while the outer operation can be viewed as a form of scaling.
- A real-valued vector space is a set $V = (\mathcal{V}, +, .)$ with operations $+, .$ such that $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and $. : R \times \mathcal{V} \to \mathcal{V}$ where $(\mathcal{V}, +)$ is an Abelian group, and the following properties hold:
    - Distributivity
    - associativity with respect to the outer operation.
    - there exists a neutral or identity element with respect to the outer operation.

Let us look at the properties of a Vector Space more carefully:

- Distributivity: $\forall \lambda \in R, \mathbf{x}, \mathbf{y} \in \mathcal{V}, \lambda.(\mathbf{x} + \mathbf{y}) = \lambda.\mathbf{x} + \lambda.\mathbf{y}$ and $\forall \lambda, \psi \in R, (\lambda + \psi).\mathbf{x} = \lambda.\mathbf{x} + \psi.\mathbf{x}$
- Associativity with respect to the outer operation: $\forall \lambda, \psi \in R, \mathbf{x} \in \mathcal{V}, \lambda.(\psi.\mathbf{x}) = (\lambda\psi).\mathbf{x}$
- Neutral element with respect to the outer operation: $\forall \mathbf{x} \in \mathcal{V}$, $1.\mathbf{x} = \mathbf{x}$
- The elements $\mathbf{x} \in \mathcal{V}$ are called vectors.
- The neutral element with respect to $(\mathcal{V}, +)$ is the zero vector $[0, 0, \ldots 0]^T$ and the inner operation is called vector addition.

Examples of vector spaces

Consider some examples of vector spaces, for example $\mathcal{V} = R^n$

- We can define addition:
  $\mathbf{x} + \mathbf{y} = (x_1, x_2, \dots x_n) + (y_1, y_2, \dots y_n) = (x_1 + y_1, x_2 + y_2, \dots x_n + y_n)$
- Multiplication by scalars:
  $\lambda\mathbf{x} = \lambda(x_1, x_2, \dots x_n) = (\lambda x_1, \lambda x_2, \dots \lambda x_n), \ \forall\lambda \in R, \mathbf{x} \in \mathbf{R^n}$.

Another example

What we call a vector need not be the standard column vector that we are accustomed to treating as a vector. We can think of $m \times n$ matrices as vectors and create a vector space out of them. Thus $\mathcal{V} = R^{m\ timesn}$ with addition and multiplication defined as below:

- Addition $\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} \ldots a_{1n} + b_{1n} \\ \vdots \\ a_{n1} + b_{n1} \ldots a_{nn} + b_{nn} \end{bmatrix}$ is defined element-wise for two matrices $A, B \in \mathcal{V}$.
- Multiplication by scalars: $\lambda\mathbf{A} = \begin{bmatrix} \lambda a_{11} \ldots \lambda a_{1n} \\ \vdots \\ \lambda a_{n1} \ldots \lambda a_{nn} \end{bmatrix}$

Vector subspaces

- Let $V = (\mathcal{V}, +, .)$ be a vector space and let $\mathcal{U} \subseteq \mathcal{V}, \mathcal{U} \neq \Phi$. then $U = (\mathcal{U}, +, .)$ is called a vector subspace of V if U is a vector space with the vector space operations $+$ and . restricted to $\mathcal{U} \times \mathcal{U}$ and $R \times \mathcal{U}$.
- We use the notation $U \subseteq V$ to denote that U is a vector subspace of V.
- If $U \subseteq V$ and V is a vector space, then U naturally inherits many properties directly from V because they hold for all $\mathbf{x} \in V$, and in particular for all $\mathbf{x} \in U \subseteq V$. These peroperties include the Abelian group property, associativity, distributivity and the neutral element.

Establishing vector subspaces

How do we show if $(\mathcal{U}, +, .)$ is subspace of V?
We need to show that

- $\mathcal{U} \neq \Phi$
- Closure of U with respect to the outer and inner operations, i.e $\forall \lambda \in R, \forall \mathbf{x}, \in \mathcal{U}, \lambda\mathbf{x} \in \mathcal{U}$ and $\forall \mathbf{x}\ , \mathbf{y} \in \mathcal{U}, \mathbf{x} + \mathbf{y} \in \mathcal{U}$.

Examples

- Let $\mathcal{V} = R^2$ and $\mathcal{U}$ be the y-axis. . Is $U = (\mathcal{U}, +, .)$ a subspace of $V = (\mathcal{V}, +, .)$? Answer: Yes, because the addition of any two vectors on the y-axis remains on the y-axis so closure with respect to the addition operation is satisfied. Also any vector on the y-axis when scaled by any real number (including 0) will yield a vector on y-axis.
- What about when we shift the y-axis one unit to the right, i.e $\mathcal{U} = \{x = 1\}$? Answer: We no longer have a vector subspace since scaling with respect to the outer operation does not have the closure property.

More examples

- What about the subset of R^2 that represents a square around the origin , i.e $-1 \leq x \leq 1, -1 \leq y \leq 1$? This is again not a subspace.
- This subspace is of particular interest to us called the nullspace of a matrix, i.e the set of solutions to a linear system of equations $A\mathbf{x} = 0$. Why is this a vector subspace?

Linear combination and linear independence

- We know that we can stay in a vector space by adding vectors belonging to the space, and scaling them?
- We are now interested in a different question $\rightarrow$ can we come up with a set of vectors such that every vector in the vector space can be represented as a sum ~~of sum~~ of these vectors with scaling as necessary?
- The answer is yes. A set of vectors capable of representing all vectors in a vector space is called a basis.
- To explore the question of finding a basis, we need to learn the concepts of linear combination and linear independence.

Linear combination

- Consider a vector space V and a set of vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\ldots$ $\mathbf{x}_k \in V$. Then every $\mathbf{v} \in V$ that is of the form $v = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 \ldots + \lambda_k \mathbf{x}_k$ is a linear combination of the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\ldots$ $\mathbf{x}_k$.
- Note that the $\mathbf{0}$ can be written trivially as a linear combination of the given k vectors $\mathbf{x}_i$s. We are however interested in non-trivial linear combinations of vectors to get the $\mathbf{0}$-vector.

Linear (in)dependence

- Consider a vector space V and k vectors $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_k \in V$. Then if there is a non-trivial linear combination of the given vectors such that $0 = \sum_{i=1}^{i=k} \lambda_i \mathbf{x}_i$ where at least one of the λ_is is non-zero, then the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_k$ are linearly dependent.
- Put another way, if the only way we can combine vectors to get the $\mathbf{0}$-vector is by letting all the λ_i be zero, then the vectors are linearly independent.

- What does the concept of linear independence capture? If the vectors are linearly dependent we can write one of the vectors in terms of the others, so that vector is redundant. On the other hand, when the vectors are linearly independent, each vector brings something to the table which the other vectors collectively cannot replace.

Linear independence

- k vectors in a vector space are either linearly independent or linearly dependent. There is no third option.
- If at least one of the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_k$ is the 0-vector, then the vectors are linearly dependent. Why? Apply the definition to see that this is the case. We can choose a non-zero λ for the 0-vector and zero λs for all the other vectors to get the linear combination to be zero.
- If all the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_k$ are non-zero, then the vectors are linearly dependent if and only if one of them is a linear combination of the others.

Using Gaussian elimination to check for linear independence

- A practical way of checking whether a bunch of vectors are linearly independent is to fill out the columns of a matrix with the given vectors and then perform Gaussian elimination to get the row-echelon form
- The pivot columns indicate all the vectors that are linearly independent, and they are all on the left.
- The non-pivot columns can be expressed as a linear combination of the pivot columns

Example

- Let us start with the matrix

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

- Gaussian elimination of this matrix will give rise to the following row-echelon form

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

- The pivot columns are the first and the third column and we see that the second column which is a non-pivot column can be expressed as a linear combination of the pivot columns to its left, which is just twice the first column.

Basis of vector space

- Do there exist a set of vectors in a vector space which "span" the entire space?
- What we mean here is that any vector $v \in V$ can be generated as a linear combination of the vectors in question.
- Is such a set of vectors unique to the vector space?
- Are all sets capable of generating all vectors in a vector space of the same size?

Generating set and basis

- Consider a vector space $V = (\mathcal{V}, +, .)$ and a set of vectors $\mathcal{A} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_k\} \subseteq V$.
- If every vector $v \in V$ can be expressed as a linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_k$, then $\mathcal{A}$ is called a **generating set** of $\mathcal{V}$.
- The set of all linear combinations of the vectors in $\mathcal{A}$ is known as the **span** of $\mathcal{A}$.
- If $\mathcal{A}$ spans the vector space V we write $V = span[\mathcal{A}]$.

Basis as smallest generating set

- Consider a vector space $V = (\mathcal{V}, +, .)$ and $\mathcal{A} \subseteq \mathcal{V}$.
- A generating set $\mathcal{A}$ of V is called minimal if there is no smaller set $\tilde{\mathcal{A}}$ such that $\tilde{\mathcal{A}} \subset \mathcal{A} \subseteq \mathcal{V}$ that spans V.
- Every linearly independent generating set of V is minimal and is called a basis of V.

Basis is not unique

- Consider the space R^3. The canonical basis is $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$
- Another basis for R^3 is $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}$
- Yet another basis for R^3 is $\begin{bmatrix}0.5\\0.8\\0.4\end{bmatrix}, \begin{bmatrix}1.8\\0.3\\0.3\end{bmatrix}, \begin{bmatrix}-2.2\\-3.3\\1.5\end{bmatrix}$

Not a basis

- Is any linear independent set a basis?
- No, we can have a linearly independent set that has too few vectors to become a basis
- As an example, consider the following set of vectors in R^4
$$\begin{bmatrix}1\\2\\3\\4\end{bmatrix}, \begin{bmatrix}2\\-1\\0\\2\end{bmatrix}, \begin{bmatrix}1\\1\\0\\4\end{bmatrix}$$
- Why is the above not a basis?

Finding a basis

There are three steps to finding a basis for a vector space. A basis of a subspace $U = span(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \subseteq R^n$ can be found by executing the following steps:

- Write the spanning vectors as columns of a matrix $\boldsymbol{A}$.
- Determine the row-echelon form of $\boldsymbol{A}$.
- The spanning vectors associated with the pivot columns are a basis of U.

Example

- Consider the vector subspace $U \subseteq R^5$ which is spanned by the following vectors:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} -1 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix}$$

- We would like to check if the vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ constitute a basis for the subspace U.
- We therefore need to check whether the given vectors are linearly independent or not.

Example

- We need to check if the only way to get $\sum_{i=1}^{i=4} \lambda_i \boldsymbol{x}_i = 0$ is to set $\lambda_i = 0$.
- We set up the matrix $\boldsymbol{A}\boldsymbol{\lambda} = 0$ where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & -4 & 8 \\ -1 & 1 & 3 & -5 \\ -1 & 2 & 5 & -6 \\ -1 & -2 & -3 & 1 \end{bmatrix}$$

- Now we perform a series of row transformations to turn this matrix into its row-echelon form to discover its pivot columns.

Example

- We get the following row-echelon form:

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

- From the above the pivot columns are columns 1, 2 and 4. Column 3 can be expressed as 2 $\times$ column 2 - 1 $\times$ column 1, so it is not linearly independent of the other columns.

Rank - definition and some properties

- The number of linearly independent rows in a matrix is equal to the number of linearly independent columns and is known as the rank of the matrix.
- Rank of a matrix $\boldsymbol{A}$ is denoted as $rk(\boldsymbol{A})$.
- The columns of $\boldsymbol{A} \in R^{m \times n}$ span a subspace $U \subseteq R^m$ with $dim(U) = rk(\boldsymbol{A})$.
- The rows of $\boldsymbol{A} \in R^{m \times n}$ span a subspace $W \subseteq R^n$ with $dim(W) = rk(\boldsymbol{A})$.

Rank - definition and some properties

- For all $\boldsymbol{A} \in R^{n \times n}$, $\boldsymbol{A}$ is invertible if and only if $rk(\boldsymbol{A}) = n$.
- For all $\boldsymbol{A} \in R^{m \times n}$, the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is solvable if and only if $rk(\boldsymbol{A}) = rk(\boldsymbol{A}|\boldsymbol{b})$, where $\boldsymbol{A}|\boldsymbol{b}$ is the augmented matrix.
- For $\boldsymbol{A} \in R^{m \times n}$, the solution space $\boldsymbol{Ax} = \boldsymbol{0}$ has dimension $n - rk(\boldsymbol{A})$. This is called the kernel or nullspace.

Row rank = column rank

- Let the number of linearly independent rows in the $m \times n$ matrix A be r, so that $rowrank(A) = r$.
- This means $\boldsymbol{A} = \boldsymbol{BC}$ where $\boldsymbol{B}$ is $m \times r$ and $\boldsymbol{C} = r \times n$ since we can express each row of $\boldsymbol{A}$ as a linear combination of r independent vectors. Here the basis vectors of the rows of $\boldsymbol{A}$ are stored in the rows of $\boldsymbol{C}$ and the combining coefficients for these rows are stored in the rows of $\boldsymbol{B}$. To get the ith row of $\boldsymbol{A}$ we use combining coefficients from the ith row of $\boldsymbol{B}$ and row vectors from the rows of $\boldsymbol{C}$.

Row rank = column rank

- We can interpret $\boldsymbol{A} = \boldsymbol{BC}$ in a different way, where the ith column of $\boldsymbol{A}$ is obtained by combining the columns of $\boldsymbol{B}$ with combining coefficients coming from the ith column of $\boldsymbol{C}$. This means that the columns of $\boldsymbol{B}$ form a generating set for the column space of $\boldsymbol{A}$. Therefore the dimension of this column space, or the column rank of $\boldsymbol{A}$ should be less than the number of columns of $\boldsymbol{B}$ which is r. Therefore we have $colrank(\boldsymbol{A}) \leq r$. But $r = rowrank(\boldsymbol{A})$, so we have $colrank(\boldsymbol{A}) \leq rowrank(\boldsymbol{A})$.
- But if we were using the matrix $\boldsymbol{A}^T$, instead of $\boldsymbol{A}$, we would have $colrank(\boldsymbol{A}^T) \leq rowrank(\boldsymbol{A}^T)$. Since the rows of $\boldsymbol{A}$ are the columns of $\boldsymbol{A}^T$ and vice-versa we have $rowrank(\boldsymbol{A}) \leq colrank(\boldsymbol{A})$. Combining the two inequalities, we have $rowrank(\boldsymbol{A}) = colrank(\boldsymbol{A})$.