Lecture 0b

Math Foundations Team

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Vector

- A point in $\mathbb{R}^n$ is represented by a n tuple i.e. $(x_1, \cdots, x_n)$ can be considered as a column vector $[x_1, \cdots, x_n]^T$ or a row vector $[x_1, \cdots, x_n]$.
- We can define addition:
  $\boldsymbol{x} + \boldsymbol{y} = (x_1, x_2, \ldots x_n) + (y_1, y_2, \ldots y_n) = (x_1 + y_1, x_2 + y_2, \ldots x_n + y_n)$
- Multiplication by scalars:
  $\lambda \boldsymbol{x} = \lambda(x_1, x_2, \ldots x_n) = (\lambda x_1, \lambda x_2, \ldots \lambda x_n), \forall \lambda \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^n$.

Linear Combinations

Let $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m \in \mathbb{R}^n$ be m vectors, then a linear combination of these vectors is an expression

$$c_1 \boldsymbol{v}_1 + \cdots + c_m \boldsymbol{v}_m$$

where c_is are any scalars.
Example: $2[1,3]^T + 3[1,0]^T = [5,6]^T$

Linear Independence and Linear Dependence

* Consider

$$c_1 \boldsymbol{v}_1 + \cdots + c_m \boldsymbol{v}_m = \mathbf{0}$$

  where $c_i \in \mathbb{R}$ and $v_i \in \mathbb{R}^n, \;\; i = 1, \cdots n$.

* Clearly if $c_i = 0, \;\; \forall i = 1, \cdots, n$, then the above equation will hold.

  Now, if no other c_is satisfies the above equation, we say that $\boldsymbol{v}_i$s are linearly independent.

* If there exists c_is that satisfies the above equation, and not all c_is are zero, then we say that $\boldsymbol{v}_i$s are linearly dependent.

Suppose $\boldsymbol{v}_i$ are linearly dependent $\Rightarrow$ There exists $j \in \{1, \cdots, n\}$ such that

$c_1\boldsymbol{v}_1 + \cdots + c_m\boldsymbol{v}_m = \boldsymbol{0}$ and $c_j \neq 0$.

$$\Rightarrow \boldsymbol{v}_j = \sum_{i=1, i\neq j}^{n} \frac{-c_i}{c_j}\boldsymbol{v}_i$$

That means if $\boldsymbol{v}_i$s are linearly dependent then we can express at least one of the vectors as linear combination of the remaining vectors.

Examples

Ex.1 Consider $[1,3]^T, [1,0]^T, [5,6]^T$. Then

$$2[1,3]^T + 3[1,0]^T = [5,6]^T$$

Therefore, $[1,3]^T, [1,0]^T, [5,6]^T$ form a linearly dependent set.

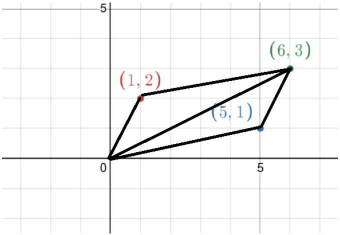
Ex.2 Consider $[1,3]^T, [1,0]^T$. Let

$$\begin{aligned} & a[1,3]^T + b[1,0]^T = [0,0]^T \\ \Rightarrow \quad & [a+b, 3a]^T = [0,0]^T \\ \Rightarrow \quad & a = 0, b = 0 \end{aligned}$$

Therefore, $[1,3]^T, [1,0]^T$ form a linearly independent set.

Ex.3 Consider $[1,2]^T, [5,1]^T, [6,3]^T$ Now,

$$[1,2]^T + [5,1]^T = [6,3]^T$$



They form a linearly dependent set.

Rank and Linear Independence

* Consider p vectors such that each of these vectors have n components. then these vectors are linearly independent if the matrix form with these vectors as row vectors has rank p.
* The are linearly dependent if rank $< p$.
* The rank r of a matrix $\boldsymbol{A}$ is also equals the maximum number of linearly independent column vectors of $\boldsymbol{A}$
  Therefore, rank$(\boldsymbol{A})$ = rank $(\boldsymbol{A}^T) = r$.
* Consider p vectors each having n components. If $n < p$ then these vectors are linearly dependent.

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 get the row-echelon form
- Pivot columns or Pivot rows are the columns or rows containing pivot elements.
- 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

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

- We get the following row-echelon form by elementary row operations

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

- 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.

Another example

Consider the vectors $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix}$

To check for linear independence we set up the equation with λs as follows: $\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \lambda_3 \mathbf{x}_3 = 0$.

As before we put the given vectors into a matrix to get:

$$\begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & -2 \\ -3 & 0 & 1 \\ 4 & 2 & 1 \end{bmatrix}$$

. Now, we can look into REF/RREF to find whether these vectors are linearly independent or not.

Dot Product

If $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors in $\mathbb{R}^n$, then $\boldsymbol{a}^T\boldsymbol{b}$ is called the dot product of $\boldsymbol{a}$ and $\boldsymbol{b}$.

Notation: $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ or $\boldsymbol{a} \cdot \boldsymbol{b}$.

$$\Rightarrow \boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^T\boldsymbol{b} = [a_1, \cdots, a_n] \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \sum_{i=1}^{n} a_i b_i$$

For example: $\boldsymbol{a} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$, then

$\boldsymbol{a} \cdot \boldsymbol{b} = (1)(-1) + (0)(2) + (2)(1) = 1.$

Properties of Dot Product

- $\langle k\boldsymbol{u} + l\boldsymbol{v}, \boldsymbol{w}\rangle = k\langle \boldsymbol{u}, \boldsymbol{w}\rangle + l\langle \boldsymbol{v}, \boldsymbol{w}\rangle, \ \forall k, l \in \mathbb{R}, \ \forall \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ (linearity)
- $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \langle \boldsymbol{v}, \boldsymbol{u}\rangle \ \forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ (symmetry)
- $\langle \boldsymbol{u}, \boldsymbol{u}\rangle \geq 0, \ \forall \boldsymbol{u} \in \mathbb{R}^n$
  $\langle \boldsymbol{u}, \boldsymbol{u}\rangle = 0$ if and only if $\boldsymbol{u} = \mathbf{0}$ (positive definite)

Norm

The length or norm of a vector $\boldsymbol{a} \in \mathbb{R}^n$ is defined as

$$\|\boldsymbol{a}\| = \sqrt{\langle \boldsymbol{a}, \boldsymbol{a} \rangle} \geq 0.$$

Properties

- $|\langle \boldsymbol{a}, \boldsymbol{b} \rangle| \leq \|\boldsymbol{a}\| \|\boldsymbol{b}\|$ Cauchy Schwarz inequality.
- $\|\boldsymbol{a} + \boldsymbol{b}\| \leq \|\boldsymbol{a}\| + \|\boldsymbol{b}\|$ Triangle inequality.

Angle between 2 vectors and orthogonality

Now we have for any $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, we have

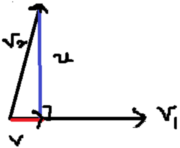
$$-1 \leq \frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\|\boldsymbol{a}\| \|\boldsymbol{b}\|} \leq 1$$

Let α be the angle between $\boldsymbol{a}, \boldsymbol{b}$. Then, we can define

$$\alpha = \cos^{-1}\left(\frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\|\boldsymbol{a}\| \|\boldsymbol{b}\|}\right)$$

Clearly, $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = 0 \Rightarrow \alpha = \frac{\pi}{2}$. Then we say $\boldsymbol{a}, \boldsymbol{b}$ are orthogonal to each other.
For example: Consider $[2,2]^T, [2,-2]^T$. Clearly $[2,2][2,-2]^T = 0$ and so they are orthogonal to each other.

Projection



The vector $\boldsymbol{v}$ is called as the projection of $\boldsymbol{v}_2$ onto the vector $\boldsymbol{v}_1$. Then $\boldsymbol{u} = \boldsymbol{v}_2 - \boldsymbol{v}$ is orthogonal to $\boldsymbol{v}_1$ and $\boldsymbol{v} = \lambda \dfrac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}$ for some scalar λ.

$$\Rightarrow \qquad (\boldsymbol{v}_2 - \boldsymbol{v})^T \boldsymbol{v}_1 = \boldsymbol{v}_2^T \boldsymbol{v}_1 - \boldsymbol{v}^T \boldsymbol{v}_1 = 0$$

$$\Rightarrow \qquad \boldsymbol{v}_2^T \boldsymbol{v}_1 = \boldsymbol{v}^T \boldsymbol{v}_1 = \lambda \|\boldsymbol{v}_1\| \;\Rightarrow \lambda = \frac{\boldsymbol{v}_2^T \boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}$$

$$\Rightarrow \qquad \boldsymbol{v} = \frac{\boldsymbol{v}_2^T \boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} = \frac{\boldsymbol{v}_2^T \boldsymbol{v}_1}{\boldsymbol{v}_1^T \boldsymbol{v}_1} \boldsymbol{v}_1$$

Basic Terminologies in Probability

- Sample Space: The set of all possible outcomes of a random experiment is known as the sample space of the experiment and is denoted by Ω.
- Events: Any subset E of sample space Ω of a random experiment is known as event.
- Algebra of Events: Union and intersection of finitely many events is an event. Complement of an event is an event.
- Mutually exclusive events: The collection of events $\{E_1, E_2, E_3, \cdots\}$ is said to be mutually exclusive, if $E_i \cap E_j = \phi$, for all $i \neq j$.
- Exhaustive Events: The collection of events $\{A_1, A_2, \cdots A_n\}$ is said to be exhaustive if $\cup_{i=1}^{n} A_i = \Omega$, where Ω is the sample space of the experiment.

Definition of Probability

Let Ω be any sample space and B be an algebra of events. A set function $P : B \to R$ is said to be a probability function if it satisfies the following three axioms:

- Axiom 1: $P(E) \geq 0$ for all $E \in B$.
- Axiom 2: $P(\Omega) = 1$
- Axiom 3: For any sequence of mutually exclusive events $E_1, E_2, \cdots$,

$$P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

Random Variable

- Random variable is a real-valued function from a sample space Ω. We use uppercase letters to denote a random variable and lowercase letter to denote the numerical values observed by random variable (rv).
- A random variable is discrete if it can assume at most a finite or countably infinite numbers of possible values.
- Example: Number of heads when a coin is tossed thrice.
- A random variable is continuous if it assumes any value in some interval or intervals of real numbers and the
- Example: The lifetime of a transistor probability that it assumes any specific value is 0.

PMF/PDF

- The probability mass function(pmf) of a discrete random variable is defined for every x by

$$p(x) = P(X = x) = P(\{s \in \Omega : X(s) = x\}) \geq 0$$
$$\sum_x p(x) = 1$$

where Ω is the sample space.

- Let X be continuous random variable. A real valued integrable function f on $\mathbb{R}$ such that

$$f(x) \geq 0, \ \forall x \in \mathbb{R}$$
$$\int_{-\infty}^{\infty} f(x)dx = 1,$$
$$P[a \leq X \leq b] = \int_a^b f(x)dx \text{ for } a, b \in \mathbb{R}$$

is called a density function for X.

Expectation or mean

- Measure of central tendency provides a very convenient way of describing a set of scores with a single number that describes the PERFORMANCE of the group. Mean is one such measure.
- The expectation or expected value of X, denoted by $E[X]$, is defined by

$$\mu = E[X] = \sum_x xp(x) \text{ or } \int_{-\infty}^{\infty} xf(x)dx$$

- Empirically, based on observations, mean is computed as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where $x_i, i = 1, \cdots, n$ are the n observations.

Variance

- Most common and most important measure of variability is the standard deviation.
- The variance of X is defined by

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x) \text{ or } \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

- Empirically, based on observations, variance is computed as

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

- The standard deviation is defined as the squareroot of variance.

Covariance

- In bivariate and multivariate case, the statistical technique used for studying the existence of relationship between all variables is through Covariance.
- If X and Y are two random variables, the covariance between X and Y is defined as

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

Therefore, empirically, it is computed as

$$\text{cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

where the ordered pair $(x_i, y_i) i = 1, \cdots, n$ are the n observations.