# Computing for Data Sciences

# Lecture 2

**Distance** is a numerical description of how far apart objects are. In physics or everyday usage, distance may refer to a physical length, or estimation based on other criteria (e.g. "two counties over"). In mathematics, a distance function or metric is a generalization of the concept of physical distance. A **metric** is a function that behaves according to a specific set of rules, and is a concrete way of describing what it means for elements of some space to be "close to" or "far away from" each other. A generalised formula for distance was hypothesised in the last class and then proven to be acceptable as a metric. Given by - $D_p(x,y) = [\sum_{i=1}^{n} |x_i - y_i|^p]^{1/p}$

in the generalized form, its properties are:

- Positive: Dist(x,y)>=0, for all real x,y
- Non-Degeneracy: Dist(x,y)=0 iff x=y
- Symmetry: Dist(x,y) = Dist(y,x) for all real x,y
- Triangle Inequality: Dist(x,z) <= Dist (x,y) + Dist (y,z) for all real x,y,z

This distance formula holds true for all values of p >=1 (at p<1, triangle inequality won't hold true). The space where the distance metric holds good is called a **metric space**. It is a collection of objects with an associated rule, or function that determines "distance" between the two objects in the space. The distance (D) between two points is defined as the edge connecting the tips of two n-dimensional vectors x and y.

$$D_p(x,y) = ||x-y||_p = [\sum_{i=1}^{n} |x_i - y_i|^p]^{1/p}$$

Where D is the distance between x and y and p is the norm

For example, p=2 gives Euclidean distance

$$||x-y||_2 = [\sum_{i=1}^{n} |x_i - y_i|^2]^{1/2}$$

At p = 2, this formula represents the distance between two points in the Euclidean space.

**Banach space** is a generic space where a metric can be defined.

**Hilbert Space** is a specific case of the Banach space where the Euclidean norm (p=2) holds.

## Norm

A norm on a vector space V is a function || · ||: **X → R** satisfying the properties-

- **Positive:** ||x|| ≥ 0 for all x in **X**
- **Non-degeneracy:** ||x|| =0 is only true if x=0
- **Scalability:** ||cv|| = |c|||x|| for all scalars c and x ∈ **X**
- **Triangle Inequality:** ||x + y|| ≤ ||x|| + ||y|| for all x and y in **X**

A vector space on which a norm can be defined is called a normed vector space. The notion of norm in a space implies the notion of distance and vice versa.

**Lp Norms**

Consider a vector $x(x_1, x_2, \ldots, x_n)$ in vector space $R^n$. Lp spaces are class of Banach spaces, defined using a generalization of the p-norm for finite dimensional spaces. The norm of x is given as-

$||x||_p = [\sum_{i=1}^{n} |x_i|^p]^{1/p}$; where p>= 1

For p=1, we get the **Manhattan (taxicab)** norm. This metric was put forward by Hermann Minkowski and the name 'taxicab' alludes to the distance a taxi has to cover in a rectangular street grid in order to reach from the origin to the destination x.

$||x||_1 = \sum_{i=1}^{n} |x_i|$

For p=2, we get the L2- norm also called the **Euclidean norm**. This norm gives the conventional distance measure of a point from the origin (that would follow from Pythagoras' theorem). This norm also holds in a vector space of complex numbers. Standard deviation is defined in this norm.

$||x||_2 = [\sum_{i=1}^{n} |x_i|^2]^{1/2}$

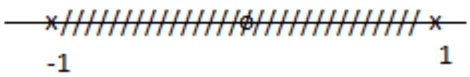
For p => ∞, we get the $L^{\infty}$ norm. This norm is equivalent to selecting the dimension with the highest weight in the n-dimensional space.

$||x||_{\infty} = max(x_1, x_2, \ldots, x_n)$

**Norms through the point of view of a Unit Circle**

A circle is set of all points that are equidistant from a given point (centre). The representation of a circle varies with the norm-space we are in.

For a vector in a 1-dimensional space: the representation of a unit circle is similar in all the norms- the region between the points -1 to 1 with centre at zero.

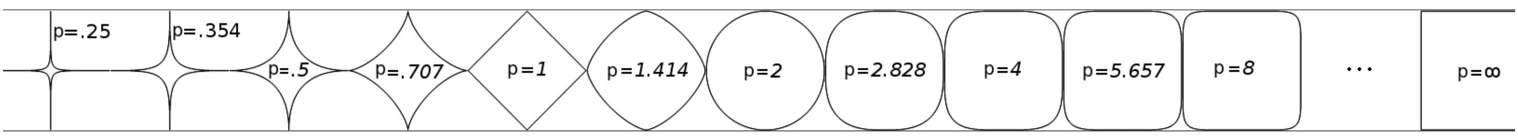



Representation of a unit circle in $\mathbb{R}^2$ space-

Using L1 norm is given by the equation |x|+|y|=1

Using L2 norm is given by $|x|^2 + |y|^2 = 1$

Using L∞ space is given by $max(|x|, |y|) = 1$




*Representation of a unit circle (2 D) in different norms*

## Some properties of matrix transformations

We can look at the operations on vectors in the form of a product of an input vector and a transformation matrix to produce an output vector.

Ideally, a list of operations that we'd want to be able to do using matrices would be:

- Arithmetic operations: Add/Subtract/Multiply (x,y)
- Amplification/Scaling Down: (k,x)
- Functions, like:
    - Negation (Inversion)
    - Rotation (By an angle)

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

$$R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

    - Reflection (About a point/line)

**To flip over the:**

| x-axis | y-axis | line y = x |
|---|---|---|

**Multiply by:**

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

    - Projection (Dropping/Adding dimensions)
    - Translation

In solving linear algebra problems, matrices allow us to view the problem from a different perspective by looking at the column view of an equation, ie. if $ax_1 + by_1 = z_1$ and $cx_2 + dy_2 = z_2$ are the two equations to be solved, they can be looked at in a 2x2 abcd matrix multiplied with an 2x1 xy matrix giving a 2x1 z matrix or it can be looked at in the form of a 2x1 ac matrix (column) multiplied with x and a 2x1 bd matrix (column) multiplied by y giving a 2x1 z matrix.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times \begin{bmatrix} x1 & x2 \\ y1 & y2 \end{bmatrix} = \begin{bmatrix} z1 \\ z2 \end{bmatrix}$$

$$\begin{bmatrix} a \\ c \end{bmatrix} \times \begin{bmatrix} x1 & x2 \end{bmatrix} + \begin{bmatrix} b \\ d \end{bmatrix} \times \begin{bmatrix} y1 & y2 \end{bmatrix} = \begin{bmatrix} z1 \\ z2 \end{bmatrix}$$

**Notion of Matrix Multiplication as Linear Transformation**

Consider a function operation on a vector **X** where X belongs to a vector space $\mathbb{R}^n$. Now if a matrix of dimension m x n is multiplied with X we get another vector of dimension m X 1. This operation is equivalent to applying a linear transformation $\mathbb{T}$on the vector such that $\mathbb{T} : \mathbb{R}^n \to \mathbb{R}^m$

Thus, we can say that the matrix multiplication is equivalent to performing a linear transformation in case of vectors. A linear transformation can also be thought of as a vector function that maps an input space (domain) to an output space (range). This equivalence is summarized below:

$$\begin{aligned}
\text{function } f:\mathbb{R}\to\mathbb{R} \;&\Leftrightarrow\; \begin{array}{l}\text{linear transformation } T_A:\mathbb{R}^n\to\mathbb{R}^m\\ \text{represented by the matrix } A\in\mathbb{R}^{m\times n}\end{array}\\
\text{input } x\in\mathbb{R} \;&\Leftrightarrow\; \text{input } \vec{x}\in\mathbb{R}^n\\
\text{output } f(x)\in\mathbb{R} \;&\Leftrightarrow\; \text{output } T_A(\vec{x})\equiv A\vec{x}\in\mathbb{R}^m\\
g\circ f\,(x)=g(f(x)) \;&\Leftrightarrow\; T_B(T_A(\vec{x}))\equiv BA\vec{x}\\
\text{function inverse } f^{-1} \;&\Leftrightarrow\; \text{matrix inverse } A^{-1}\\
\text{zeros of } f \;&\Leftrightarrow\; \text{null space of } T_A\equiv\text{null space of } A\equiv\mathcal{N}(A)\\
\text{image of } f \;&\Leftrightarrow\; \text{image of } T_A\equiv\text{column space of } A\equiv\mathcal{C}(A)
\end{aligned}$$

*(Source: No Bullshit Guide to Linear Algebra by Ivan S)*

**Geometrical Interpretation of Determinant**

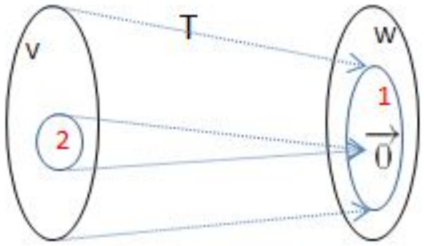
If the rows of a matrix can be thought to form the edges of a geometrical shape, then the determinant is the volume occupied by this shape. If two rows are linearly dependent then they differ only by a scalar factor. This implies that the space spanned by the matrix reduces by one (independent) dimension. Since the volume of a (n-1) dimensional cube in an n-dimensional space is zero; the determinant of such a matrix would also be zero.

If the determinant of a matrix is zero, it cannot be invertible.

The number of linearly independent rows in a matrix is given by the **rank** of a matrix. Thus, a matrix can be invertible only if its rank= row-size.

**Discussion on invertibility in case of matrices and linear transformations**

A linear transformation is a mapping from an input space V to an output space W. Here 1 denotes the column space of the matrix and 2 denotes the null space.



1. Column Space:

Let K be a field of scalars. Let A be an m × n matrix, with column vectors $v_1, v_2, ..., v_n$. A linear combination of these vectors is any vector of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n,$$

where $c_1, c_2, ..., c_n$ are scalars. The set of all possible linear combinations of $v_1, ... ,v_n$ is called the column space of A. That is, the column space of A is the span of the vectors $v_1, ... , v_n$.

Any linear combination of the column vectors of a matrix A can be written as the product of A with a column vector:

$$A\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} c_1a_{11}+ & \cdots & +c_na_{1n} \\ \vdots & \vdots & \vdots \\ c_1a_{m1}+ & \cdots & +c_na_{mn} \end{bmatrix} = c_1\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + c_n\begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}$$
$$= c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

Therefore, the column space of A consists of all possible products Ax, for $x \in C^n$. This is the same as the image (or range) of the corresponding matrix transformation.

2. Null Space (Kernel):

Consider a linear map represented as a m × n matrix A with coefficients in a field K (typically the field of the real numbers or of the complex numbers) and operating on column vectors x with n components over K. The kernel of this linear map is the set of solutions to the equation $A\vec{x} = \vec{0}$, where 0 is understood as the zero vector. The dimension of the kernel of A is called the nullity of A. In set-builder notation,

$$N(A) = Null(A) = ker(A) = x \epsilon K^n | Ax = 0$$

The matrix equation is equivalent to a homogeneous system of linear equations:

$$A\mathbf{x} = \mathbf{0} \quad \Leftrightarrow \quad \begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n & = 0 \\ a_{21}x_1 & + & a_{22}x_2 & + \cdots + & a_{2n}x_n & = 0 \\ \vdots & & \vdots & & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + \cdots + & a_{mn}x_n & = 0. \end{array}$$

Thus the kernel of A is the same as the solution set to the above homogeneous equations.

A linear transformation T whose column space is equal to the co-domain is called **surjective or onto**. An onto relation signifies that all possible outputs can be achieved through some input. If only some of the output space is covered, it is an into transformation.

A linear transformation in which $\vec{w}(x) = \vec{w}(y)$ implies that x=y is said to **be injective or one-one**. If x!=y, then the transformation is termed as many-one.

If a transformation is both surjective and injective, then it is termed as bijective. Bijection is a necessary condition for invertibility. This is because one-to-one correspondence between input space and output space and vice-versa is only possible in case of bijectivity. Thus, the size of the input space and output space must be the same. Therefore, by the notion of equivalence between linear transformation and matrix operation on vectors discussed in previous sections, only square matrices (n x n) can be invertible.