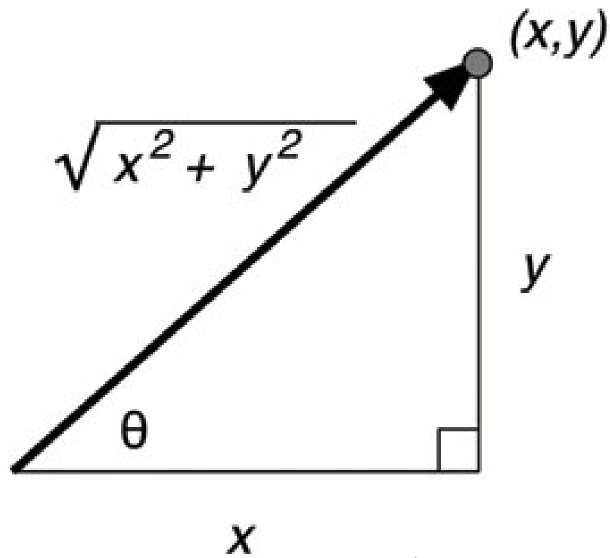
Why Learn Linear Algebra?

As we all know that as humans we are resricted to only perceiving data in three dimensions. But most ML techniques deal with high dimensional data and hence it is not possible to visualise and analyse the data in plots or through statistics techniques, thats where linear algebra kicks in, to help us analyse the data mathematically and visualise data using matrices.Also Many machine learning concepts are tied to linear algebra. For example, PCA requires eigenvalues and regression requires matrix multiplication.

Concepts Discussed

Vector Algebra

1.Vector - a quantity having direction as well as magnitude, especially as determining the position of one point in space relative to another.

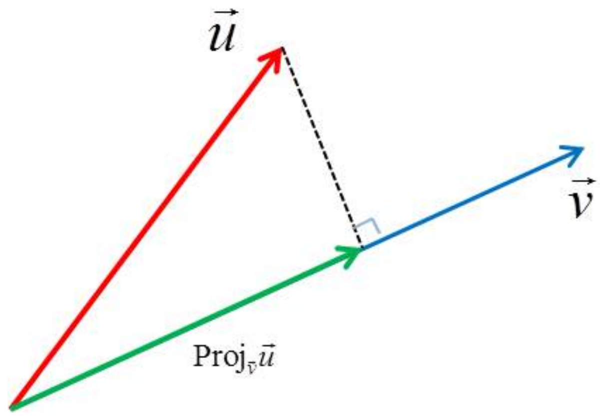


$$Magnitude(||a||) = \sqrt{x^2 + y^2}$$

2.Unit Vector - a vector having magnitude = 1. Any vector can be made a unit vector by dividing the vector with its magnitude. Converting a vector into unit vector is also called vector normalization.

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{|\mathbf{v}|}$$

3.Projcetions - The vector projection of a vector a on (or onto) a nonzero vector b (also known as the vector component or vector resolution of a in the direction of b) is the orthogonal projection of a onto a straight line parallel to b. (In the diagram below u's projection on v is shown in a green line)



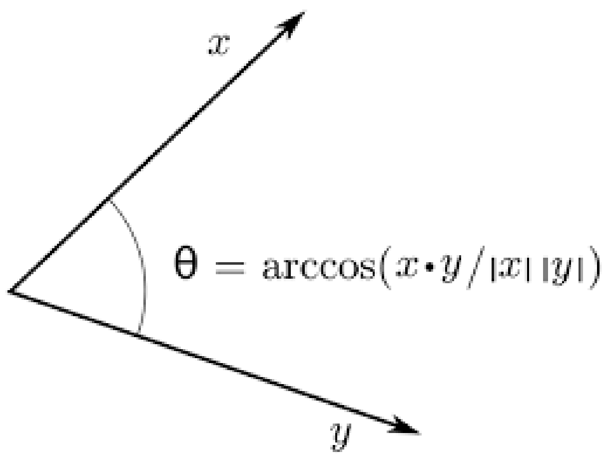
It is a vector parallel to b, defined as

$$\mathbf{a}_1 = a_1 \hat{\mathbf{b}}$$

where a1 is a scalar, called the scalar projection of a onto b, and it is multiplied with the unit vector in the direction of b. In turn, the scalar projection is defined as

$$a_1 = |\mathbf{a}| \cos\theta = \mathbf{a} \cdot \hat{\mathbf{b}} = \mathbf{a} \cdot \frac{\mathbf{b}}{|\mathbf{b}|}$$

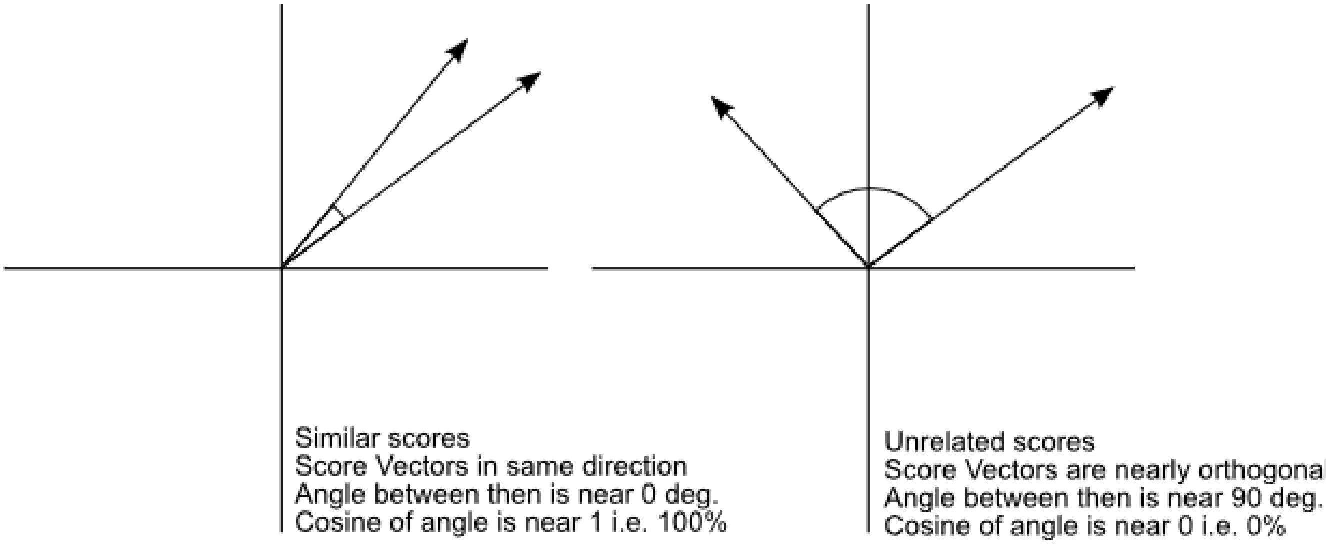
4.Dot Product - of two vectors a and b is a scalar quantity equal to the product of magnitudes of vectors multiplied by the cosine of the angle (α) between vectors: a · b = |a| · |b| cos α
It is also equal to the sum of pairwise products of coordinate vectors a and b : a · b = ax · bx + ay · by
It can also be used to calculate the angle between the vectors as so:



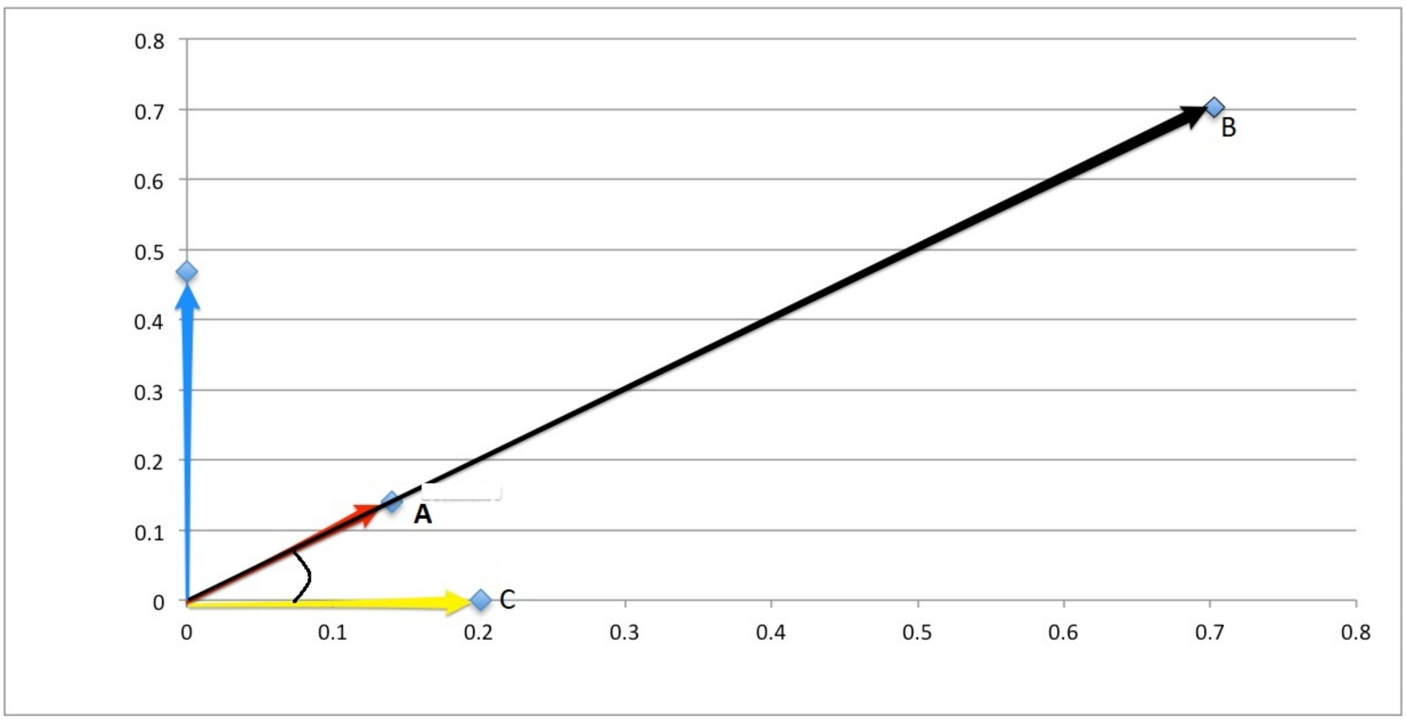
5.Cosine Similarity - Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

<u>Example</u>:



<u>Cosine Similarity Fails in the below case</u>:



As we can see above even though A & B are far apart and may have no similarity, if we use cosine similarity to compute similarity between A & B we would get that they are similar as cosine similarity does not take into account the distance between points only the angle between is measured and between A & B the angle is 0 degree hence cosine of 0 = 1 therefore they are assumed to be similar by cosine simlarity even when they are not.

6.Coordinate Geometry

6.1.Change of Axes during Normalization - $x^{'} = (x - \mu)/\sigma$.

On performing the operation x- μ we are shifting the axes of the distibution of points such that the mean of the distribution of points lies at the origin.

On dividing the above result obtained with σ we are sqeezing or expanding the points such that their variance becomes equal to 1.

6.2.Data-Point as a Line

Suppose our data consists of 3 data-points (3 rows) and each data point has three features (3 columns) then our data matrix will be as below where the superscript denotes the data-set number and subscript denotes the feature number. so f_2^1 means third feature of the second dataset.

$$\begin{pmatrix} f_0^0 & f_1^0 & f_2^0 \\ f_0^1 & f_1^1 & f_2^1 \\ f_0^2 & f_1^1 & f_2^2 \end{pmatrix}$$

Considering each data-set (row) as a vector we get $f_1x_1 + f_2x_2 + f_0x_0$ comparing this with the equation of line $ax + by + c = 0$. The same can be written as: $\begin{pmatrix} f_1 & f_2 & f_0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = 0$

We know vectors are n-dimensional column vectors therefore $f^T.x = 0$ Now as dot product of the above two are zero hence f^T is perpendicular to direction of line x (since cos90 = 0)

References

Link to the MIT OCW course on Linear Algebra: https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/video-lectures/ (https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/video-lectures/)