1 Learning Outcomes

- Able to write and interpret equations written in vector notation
- Able to convert equations between vector and scalar notation
- Differentiate between row and column vectors
- Calculate transposes, dot products, matrix-vector multiplication, and matrix-matrix multiplication
- Determine whether a given dot product, matrix-vector multiplication, and matrix-matrix multiplication operation is valid for two numerical objects (e.g., vectors or matrices)
- Explain why linear algebra is relevant to machine learning
- Define the equation for a hyperplane

2 Overview

These notes provide an overview to the basics of linear algebra that is needed in this course. We will introduce the basic mathematical objects: vectors and matrices, and their corresponding arithmetic operations.

In machine learning, vectors and matrices can be seen as data structures or data containers that collect measurements and help make mathematical notations more concise and compact. Many of the high-level languages like Python support the use of vector notations which make model fitting and prediction more efficient (through parallelization). Moreover, certain machine learning problems are set up and solved using tools from linear algebra (for example: closed-form solutions to Linear regression and dimensionality reduction techniques like Principal Component analysis: PCA). Linear algebra provide a useful tool to interpret certain machine learning concepts geometrically. Therefore having a good understanding of linear algebra is important to gaining a good understanding and to working with many machine learning algorithms.

3 Vector

Before we define vectors, let's review what scalars mean. A scalar is a single number that represents a numerical quantity: age, weight, temperature. We denote a scalar by a lower-case letter, for instance: a, s. The notation $a \in \mathbb{R}$ means that a is a real-valued scalar, where $\mathbb{R}$ represents the set of all real-valued scalars.

A vector is an array of numerical values. It collects an ordered list of numbers to represent one entity or one mathematical object. A vector is usually represented as a vertical array surrounded by square or curved brackets, or as numbers separated by commas and surrounded by parentheses:

$$\begin{bmatrix} 5 \\ 2 \\ 7 \end{bmatrix}, \begin{pmatrix} 5 \\ 2 \\ 7 \end{pmatrix}, (5, 2, 7)$$

Each value in a vector is called an element, entry or component. The number of elements in a vector is the length or size of the vectors. We denote a vector by a lower-case bold letter (for example $\mathbf{x}$, $\mathbf{y}$) and its i^{th} element by x_i, where the subscript i is an integer index that runs from 1 to n, where n is the size of the vector:

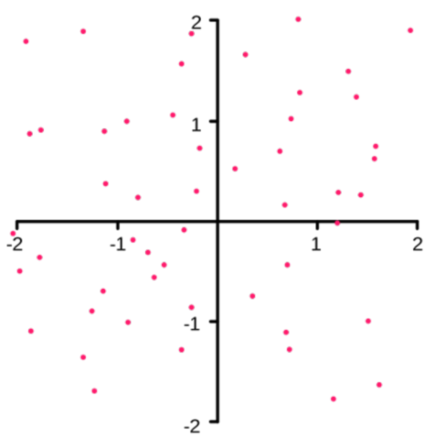
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

We refer to vectors with real components as real vectors. The notation $\mathbf{x} \in \mathbb{R}^n$ means that a $\mathbf{x}$ is vector with n real entries, and $\mathbb{R}^n$ is the set of all real vectors with size n.

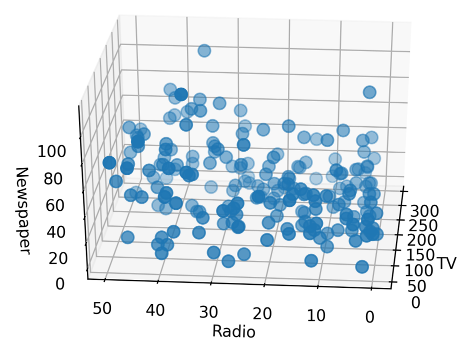
In machine learning, one example where we use vectors is when collect features to represent one data sample or example. For instance, if we were training a model to predict the sale price of houses in a given area, we might associate each house with a vector whose components correspond to its number of bedrooms, area in square foot, number of bathrooms, zip code, etc. This vector of features is what represents each house in the data. If we were studying the presence of heart disease in patients, we might represent each patient by a vector whose components correspond to their age, resting blood pressure, cholesterol level, fasting blood sugar and maximum heart rate achieved. We can also group the labels of the training data in a vector: vector of labels, or we can also group the parameters of a model in a vector: vector of parameters.

3.1 Geometry of vectors

A vector can be interpreted geometrically as a point in space or a geometric vector: an object with direction and magnitude.



(a) Examples of 2D-points



(b) Examples of 3D-points

Figure 1: Examples of points in two and three dimensional spaces

Point: The components of a n-vector can be seen as the coordinates of a point in the n-dimensional space. The coordinates determine the position of the point in the space, and two points are different if they have different positions in the space. The figures 1a and 1b show examples of points in two and three dimensional spaces. Figure 1b show the advertising budgets in 200 different markets for different media channels: TV, radio and newspaper. Each point corresponds to a single market and shows its advertising budgets for TV, radio and newspapers.

Geometric Vector: A n-vector can be also interpreted as an object that has magnitude and direction. It can be depicted as a directed line segment going from one point to another, where the components of the vector tell us in which direction to go and how far. For example, the vector $(2, 3)$ can be represented as a line going from the origin to the point $(2, 3)$. To geometrically represent the vector $(2, 3)$ we do not need to always start from the origin. We can start from any point and then take 2 steps to the right and 3 steps up. Any of the vector drawn below can represent the vector $(2, 3)$.

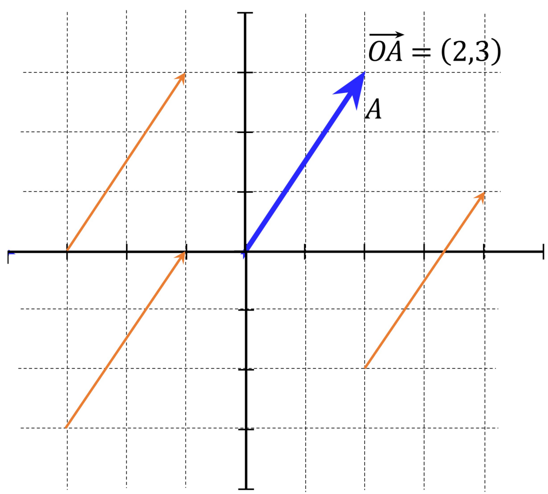


Figure 2: A vector can be visualized as a directed line segment. In this case, every vector drawn is a representation of the vector $(2, 3)$.

Given the position of two points, we can compute the components of the geometric vector going from one point to another. For instance, the components of the geometric vector represented in figure 3 is given by:

$$\begin{bmatrix} 2-(-2) \\ 3-(-1) \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

In machine learning, when we want interpret a model geometrically, we might treat a given vector as a point or as a geometric vector depending on the model.

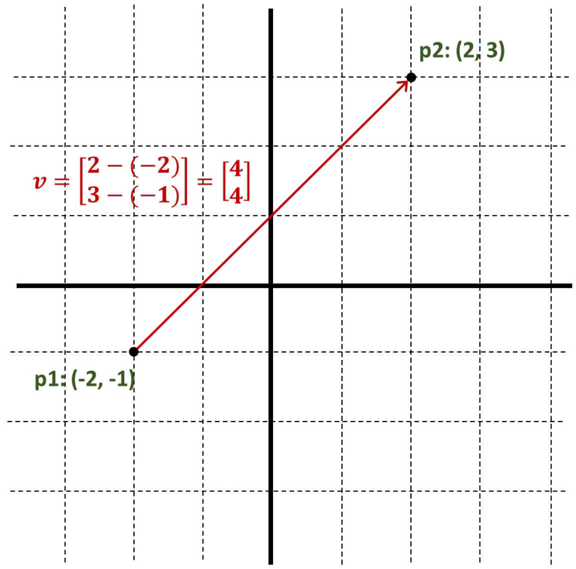


Figure 3: The components of the vector going from point p_1 to p_2

3.2 Addition of two vectors and Scalar Multiplication

- If $\mathbf{x} \in R^n$ and $\mathbf{y} \in R^n$, then their vector addition $\mathbf{x} + \mathbf{y}$ is defined as:

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

- If $\mathbf{x} \in R^n$ and $\lambda \in \mathbb{R}$ (scalar), then scalar multiplication is defined as:

$$\lambda \mathbf{x} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{bmatrix}$$

3.3 Norm and Dot Product

Dot Product

The dot product is one of the most fundamental operations in linear algebra. Given two real n-vectors $\mathbf{x}$ and $\mathbf{y}$, their dot product is the scalar given by the sum of their element-wise products. The dot product is denoted by $\mathbf{x} \cdot \mathbf{y}$ and defined as:

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

The dot product can be also denoted: $\mathbf{x}^T \mathbf{y}$ where the notation T means transpose and will be explained in the next section. One example in machine learning that heavily uses dot products is linear models. One step in building linear models involves the computation of a weighted sum of the features: if $\mathbf{x}$ represents the vector of features and $\mathbf{w}$ represents the vector of weights, then the weighted sum of the features is given by: $\sum_{i=1}^{n} w_i x_i$ which can be equivalently expressed as the dot product: $\mathbf{x}.\mathbf{w} = \mathbf{x}^T\mathbf{w}$.

Norm

The magnitude norm (Euclidean norm or L2-norm) of an n-vector $\mathbf{x}$ denoted by $\|\mathbf{x}\|$, is the square root of the sum of the squares of its elements,

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Given the definition of the dot product, it is easy to see that:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{\mathbf{x}^T\mathbf{x}}$$

Unit vector: a unit vector is vector with norm of 1. Given a vector $\mathbf{x}$, its corresponding unit vector $\mathbf{x}_u$ is given by:

$$\mathbf{x}_u = \frac{\mathbf{x}}{\|\mathbf{x}\|}$$

Euclidean distance

The Euclidean distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ is the norm of their difference:

$$\mathrm{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

When the distance between two n-vectors is small, we say they are close or nearby, and when the distance is large, we say they are far. Geometrically, the distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ is the distance between the two points with coordinates $\mathbf{u}$ and $\mathbf{v}$ respectively as shown in figure 4. This can be seen by directly applying the definition of norm:

$$\mathrm{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}$$

Moreover the norm of the vector $\|\mathbf{u}\|$ is the magnitude of the line segment between the origin and the point with coordinates given by $\mathbf{u}$.

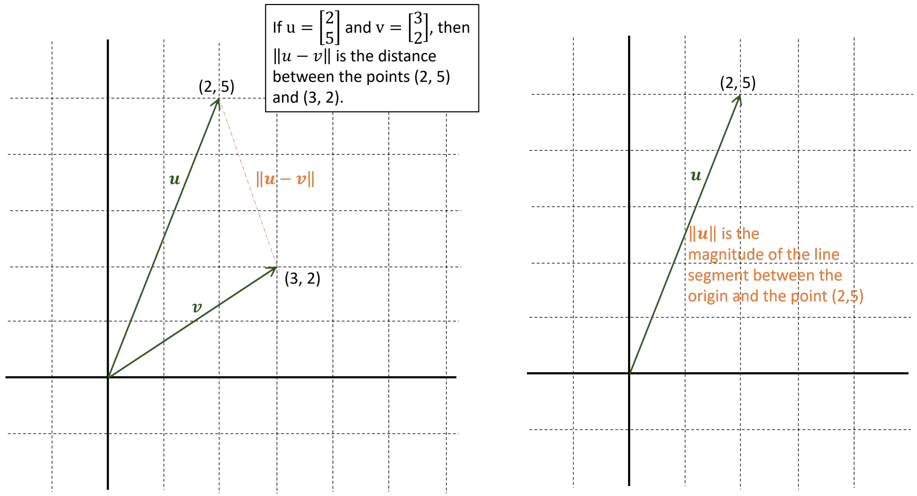


Figure 4: Euclidean distance between two vectors

Norm and dot product:
The dot product of two vectors **u** and **v** is also equal to the following:

$$\mathbf{u} \cdot \mathbf{v} \; = \|u\| * \|v\| * \cos(\alpha)$$

here α is the angle between the two vectors as depicted in the figure 5.

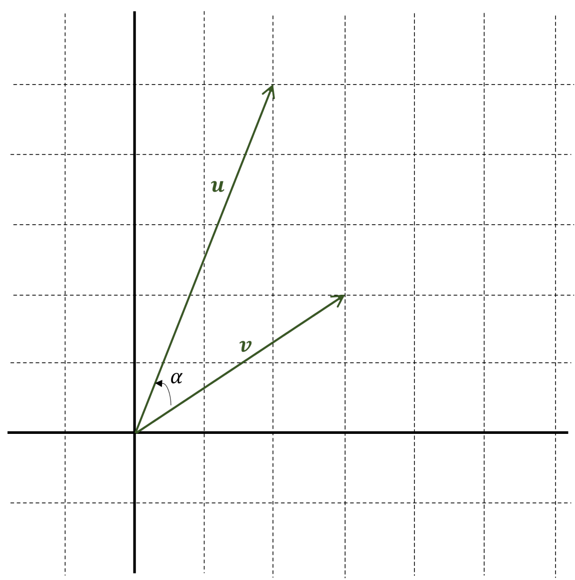


Figure 5: The angle between two vectors

Using this definition, we can see a relation between the value of α and the sign of the dot product $\mathbf{u} \cdot \mathbf{v}$:

- If $\alpha = \frac{\pi}{2}$, then $\mathbf{u} \cdot \mathbf{v} \; = 0$, the vectors are said to be orthogonal.

- If the angle is zero,then $\mathbf{u} \cdot \mathbf{v} \ = \|u\|\|v\|$, the vectors are parallel or aligned. Each vector is a positive multiple of the other.
- If $\alpha = \pi$, then $\mathbf{u} \cdot \mathbf{v} \ = -\|u\|\|v\|$, the vectors are anti-parallel or anti-aligned. Each vector is a negative multiple of the other.
- If $\alpha < \frac{\pi}{2}$, then $\mathbf{u} \cdot \mathbf{v} \ > 0$, the vectors are said to be acute.
- If $\alpha > \frac{\pi}{2}$, then $\mathbf{u} \cdot \mathbf{v} \ < 0$, the vectors are said to be obtuse.

These different cases are illustrated in figure 6.

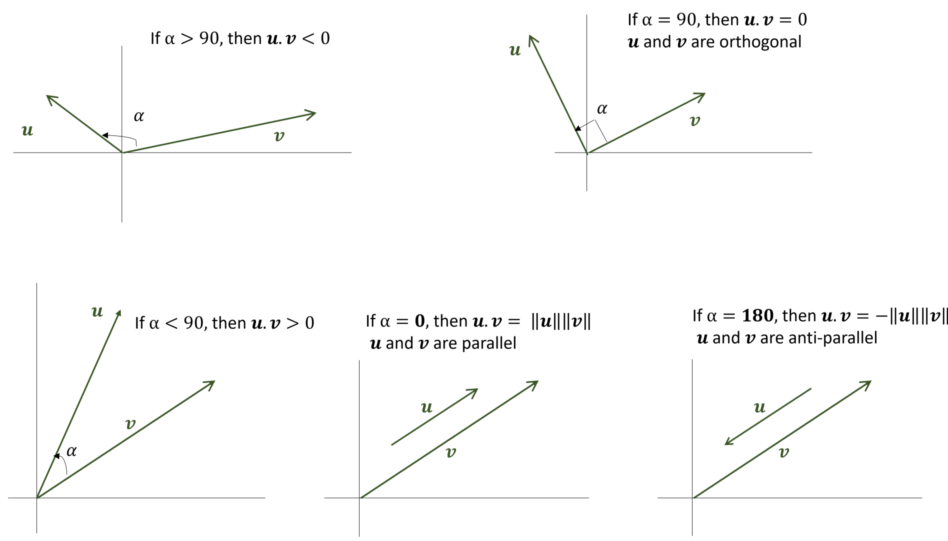


Figure 6: Signs of the dot product

4 Matrices

Just as vectors collect a set of scalars, a matrix collects an ordered list of vectors and it allows operations on groups of vectors in one formula. A matrix is a rectangular array of numbers written between squared brackets:

$$\begin{bmatrix} 3 & 5 & 7 \\ 2 & 6 & 8 \end{bmatrix}$$

A matrix is described by its number of rows and number of columns. The array above for example has 2 rows and 3 columns. We will denote a matrix by a capital letter (for example A). The notation $A \in \mathbb{R}^{n \times m}$ means that the matrix A consists of n rows and m columns of real-valued scalars. We say that A is of size $n \times m$ or it is an $n \times m$ matrix. The element of A that exists at the i^{th} row and j^{th} column is denoted by a_{ij}. The matrix $A \in \mathbb{R}^{n \times m}$ can be illustrated as follows:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

Square matrices. A square matrix is a matrix that has an equal number of rows and columns.

Matrix Transpose. If A is an $n \times m$ matrix, its transpose, denoted A^T, is the $n \times m$ matrix, where each row in A becomes a column in A^T and each column in A becomes a row in A^T. For example,

$$\begin{bmatrix} 3 & 5 & 7 \\ 2 & 6 & 8 \end{bmatrix}^T = \begin{bmatrix} 3 & 2 \\ 5 & 6 \\ 7 & 8 \end{bmatrix}$$

Row and column vectors. An n-vector can be interpreted as column vector, meaning a $n \times 1$ matrix or a row vector, meaning a $1 \times n$ matrix. In this course, we will assume that n-vector is a column vector unless stated otherwise, and we will not distinguish between vectors and matrices with one column. In other words, we will assume that the spaces $\mathbb{R}^n$ and $\mathbb{R}^{n \times 1}$ are equivalent. If $\mathbf{x}$ is n-vector, then we will assume that $\mathbf{x}$ is a column vector and that $\mathbf{x}^T$ is a row vector.

4.1 Operations on matrices

Matrix-vector multiplication: If A is an $n \times m$ matrix and $\mathbf{x}$ is an m-vector, then the matrix-vector product $A\mathbf{x}$ results in a n-vector $\mathbf{y}$ where the i^{th} element of y, y_i, is the dot product between the i^{th} row of A and the vector $\mathbf{x}$. Example:

$$\begin{bmatrix} 3 & 5 & 7 \\ 2 & 6 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 + 10 - 7 \\ 2 + 12 - 8 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

Note that for $A\mathbf{x}$ to be a valid operation, the number of elements of $\mathbf{x}$ needs to be equal to the number of columns of A.

Matrix-matrix multiplication: Two matrices can be multiplied if their dimensions are compatible: the matrices A and B can be multiplied if the number of columns of A is equal to the number of rows B. Suppose A has size $n \times p$ and B has size $p \times m$, then AB results in the $n \times m$ matrix C where c_{ij} is the dot product of the i^{th} row of A and j^{th} column of B. Example:

$$\begin{bmatrix} 3 & 5 & 7 \\ 2 & 6 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 2 & 1 & -1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 5 & -2 & 7 \\ 6 & 6 & -4 & 8 \end{bmatrix}$$

Not that the multiplication of two matrices is different then their element-wise multiplication, which is know as the Hadamard product.

Geometrical Interpretation of Matrix Multiplication: One interpretation of Matrix-vector multiplication is that it transforms a vector into a new vector. This transformation is known as linear mapping. Assume for example a vector $\mathbf{u} \in \mathbb{R}^2$ and $A \in \mathbb{R}^{2\times 2}$ and the following linear transformation that transforms a vector $\mathbf{u}$ into vector $\mathbf{v}$:

$$\mathbf{u} \xrightarrow{A} \mathbf{v} = A\mathbf{u}$$

where $\mathbf{v} \in \mathbb{R}^2$. If we assume that $\mathbf{u}$ is a point in the 2D space, then multiplying $\mathbf{u}$ by a specific A can result in rotating or reflecting $\mathbf{u}$ (depending on the choice of A).

- **Rotation**: If A takes the following form:

  $$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

  Then $\mathbf{v} = A\mathbf{u}$ is the vector obtained by rotating $\mathbf{u}$ by θ radians counterclockwise. Example: assume $\theta = \pi$. Rotating a point by π counterclockwise should result in moving the point to the opposite quadrant (negating the coordinates). Let's verify it. If $\theta = \pi$, then

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

  and

  $$\mathbf{v} = A\mathbf{u} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -u_1 \\ -u_2 \end{bmatrix}$$

  Now let's say we want to apply this particular transformation to a set of points: $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$ and $\mathbf{u}^{(3)}$. We can collect these vectors into one matrix:

  $$U = \begin{bmatrix} \mathbf{u}^{(1)} & \mathbf{u}^{(2)} & \mathbf{u}^{(3)} \end{bmatrix}$$

  and represent the collective transformation in one formula:

  $$V = AU = A \begin{bmatrix} \mathbf{u}^{(1)} & \mathbf{u}^{(2)} & \mathbf{u}^{(3)} \end{bmatrix} = \begin{bmatrix} A\mathbf{u}^{(1)} & A\mathbf{u}^{(2)} & A\mathbf{u}^{(3)} \end{bmatrix}$$

  The first column of V represents the coordinates of the vector obtained after transforming the first column vector of U (and so on).

  For instance, suppose we want to transform the three points: $(1, 3)$, $(-1, 2)$ and $(3, -1)$, by applying the linear transformation given by A for $\theta = \pi$. We can collect the points into the matrix U as follows:

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

  Then we apply the transformation:

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

  This example is illustrated in figure 7.

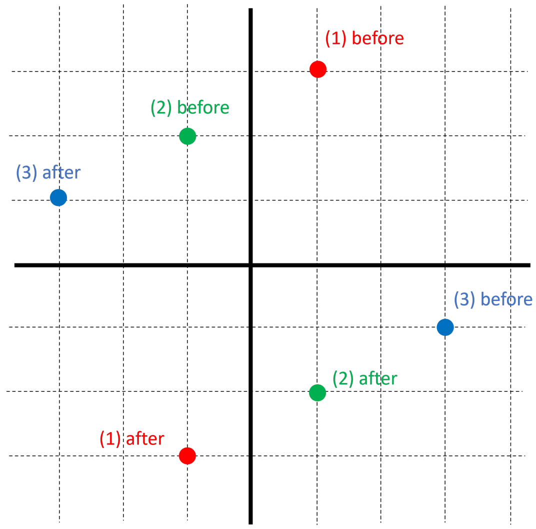


Figure 7: Example of linear transformation (rotation by π)

- **Reflection**: If A takes the following form:

$$\begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}$$

Then $\mathbf{v} = A\mathbf{u}$ is the vector obtained by reflecting $\mathbf{u}$ through the line that passes through the origin, inclined θ radians with respect to horizontal. Example: assume $\theta = \frac{\pi}{4}$. Reflecting a point through the line that passes through the origin and that is inclined by $\frac{\pi}{4}$ radians, should result in swapping the coordinates. Let's verify it. If $\theta = \frac{\pi}{4}$, then

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and

$$\mathbf{v} = A\mathbf{u} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} u_2 \\ u_1 \end{bmatrix}$$

We can also apply this transformation to a collection of vectors. For instance:

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

This example is illustrated in figure 8.

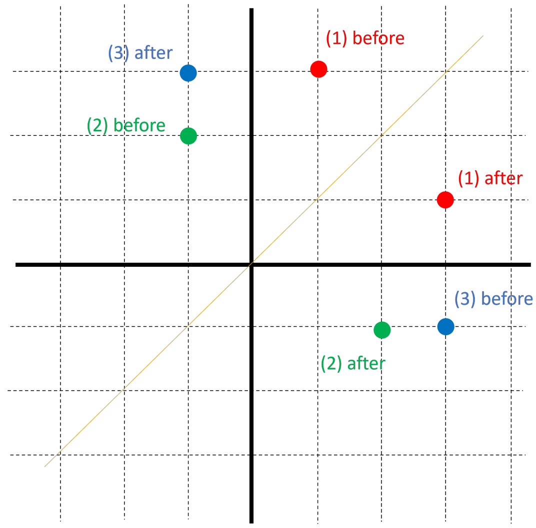


Figure 8: Example of linear transformation (reflection around the 45 degree line)

5 Hyperplane

5.1 Line

In a 2-dimensional space, we know how to plot and get the equation of a line: $y = mx + c$ where m is the slope and c is the y-intercept. We can equivalently represent the equation of the line as follows:

$$w_1 x + w_2 y + w_0 = 0$$

where w_1, w_2 and w_0 are constant that are related to the slope ($m = -\frac{w_1}{w_2}$) and y-intercept of the line ($c = -\frac{w_0}{w_2}$) . In this course, we might label the horizontal axis with a label different than x (like x_1) and the vertical axis with a label different than y (like x_2) as shown in figure 9. The equation of the line is then given by:

$$w_1 x_1 + w_2 x_2 + w_0 = 0$$

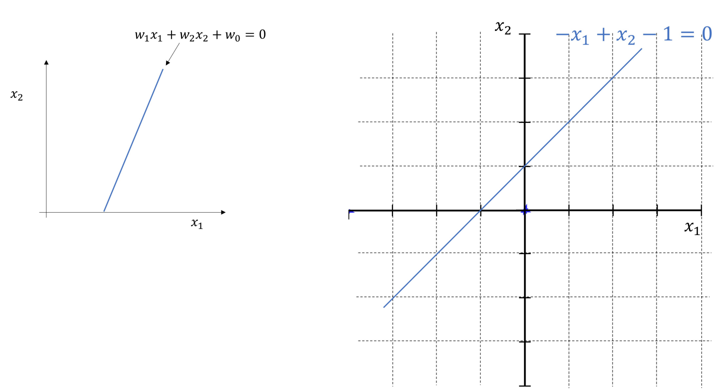


Figure 9: An example of line's equation

A line defined by the above equation represents the set of all points $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ that satisfy $w_1x_1 + w_2x_2 + w_0 = 0$. The coefficients w_1 and w_2 can be collected in a vector $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$. This vector is known as the ***normal vector*** of the line: it is the vector that is orthogonal to the line, and it is what determines the orientation of the line. If we are given the normal vector $\mathbf{w}$ and the coordinates of a particular point $\mathbf{p}$ on the line, we can determine the equation of the line.

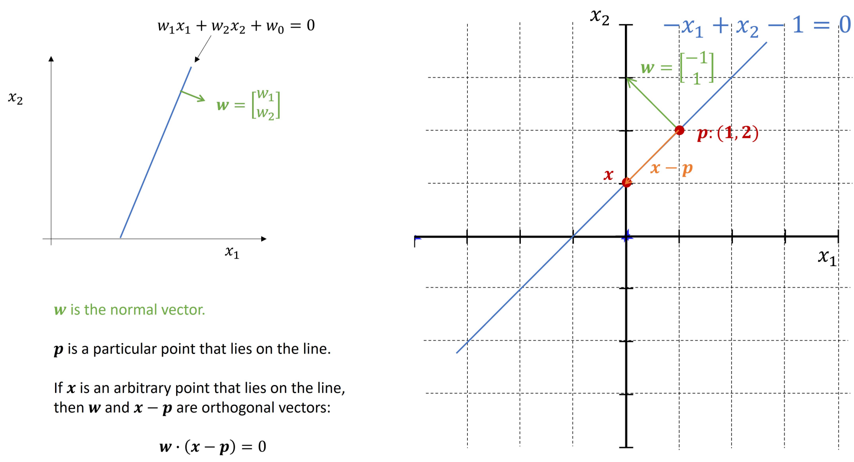


Figure 10: A line can be defined by a normal vector and a point on the line.

For example, in figure 10, the line is given by the following equation: $-x_1 + x_2 - 1 = 0$. If we don't know the equation of the line, but we are told that the line has $\mathbf{w} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ as its normal vector and that it passes through the point $\mathbf{p} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then we can retrieve its equation. How? Since $\mathbf{w}$ is a normal vector, then it is perpendicular to the line. This implies that if we take any arbitrary point $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ on the line then the two vectors $\mathbf{w}$ and $\mathbf{x} - \mathbf{p}$ must be orthogonal, i.e.,

$$\mathbf{w} \cdot (\mathbf{x} - \mathbf{p}) = 0 \implies w_1(x_1 - 1) + w_2(x_2 - 2) = 0 \implies -(x_1 - 1) + (x_2 - 2) = 0 \implies -x_1 + x_2 - 1 = 0$$

The line divides the 2-dimensional space into two half-spaces:

- a half-space defined by $w_1x_1 + w_2x_2 + w_0 > 0$ (figure 11):

  any point $\mathbf{x}$ that lies on the same side of the normal vector must satisfy $w_1x_1 + w_2x_2 + w_0 > 0$ (this is because the angle between the normal vector and $\mathbf{x} - \mathbf{p}$ is less than 90);

- a half-space defined by $w_1x_1 + w_2x_2 + w_0 < 0$ (figure 12):

  any point $\mathbf{x}$ that lies on the other side of the normal vector must satisfy $w_1x_1 + w_2x_2 + w_0 < 0$ (this is because the angle between the normal vector and $\mathbf{x} - \mathbf{p}$ is greater than 90);

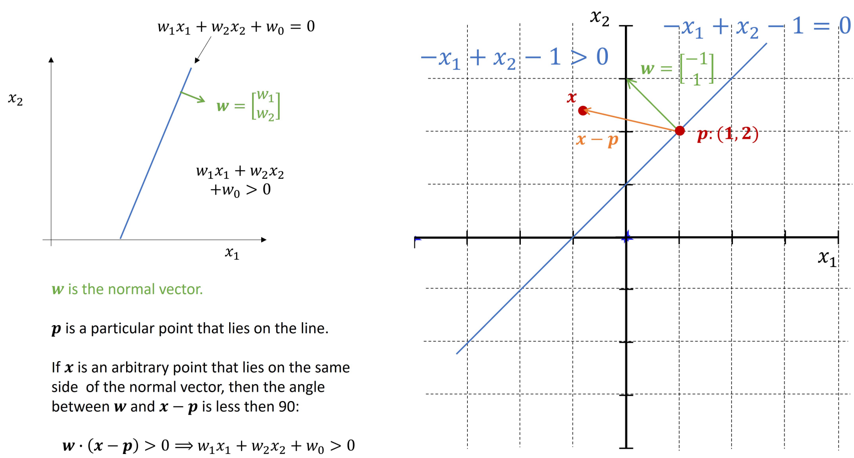


Figure 11: Any point $\mathbf{x}$ that lies on the same side of the normal vector satisfies $w_1x_1 + w_2x_2 + w_0 > 0$

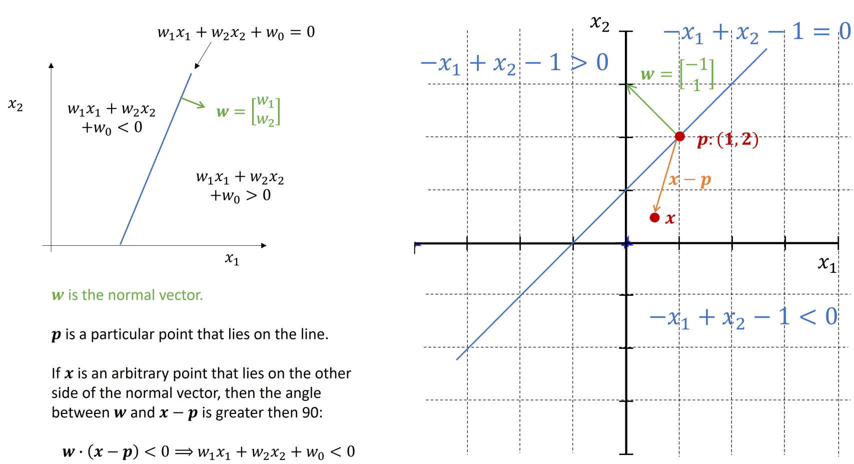


Figure 12: Any point $\mathbf{x}$ that lies on the other side of the normal vector satisfies $w_1x_1 + w_2x_2 + w_0 < 0$

5.2 Generalization of line

In a 3-dimensional space, a plane is the generalization of the line. A plane is given by:

$$w_1x_1 + w_2x_2 + w_3x_3 + w_0 = 0$$

In higher dimensional space, a hyperlane is the generalization of the line and plane. In m-dimensional space, the equation of a hyperplane is given by:

$$w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_mx_m + w_0 = 0 \implies \mathbf{w}^T\mathbf{x} + w_0 = 0$$

The same properties we mentioned when discussing the line are still valid with hyperplane:

- $\mathbf{w}$ is the normal vector of the hyperplane that defines its orientation.

- A hyperplane can be defined by a normal vector and a particular point of the hyperplane.
- Any point $\mathbf{x}$ that lies on the same side of the normal vector satisfies: $\mathbf{w}^T\mathbf{x} + w_0 > 0$.
- Any point $\mathbf{x}$ that lies on the other side of the normal vector satisfies: $\mathbf{w}^T\mathbf{x} + w_0 < 0$

6 References

- Geometry of vectors
- Vectors
- Introduction to Linear Algebra for Applied Machine Learning with Python - Introductory level linear algebra notes for applied machine learning