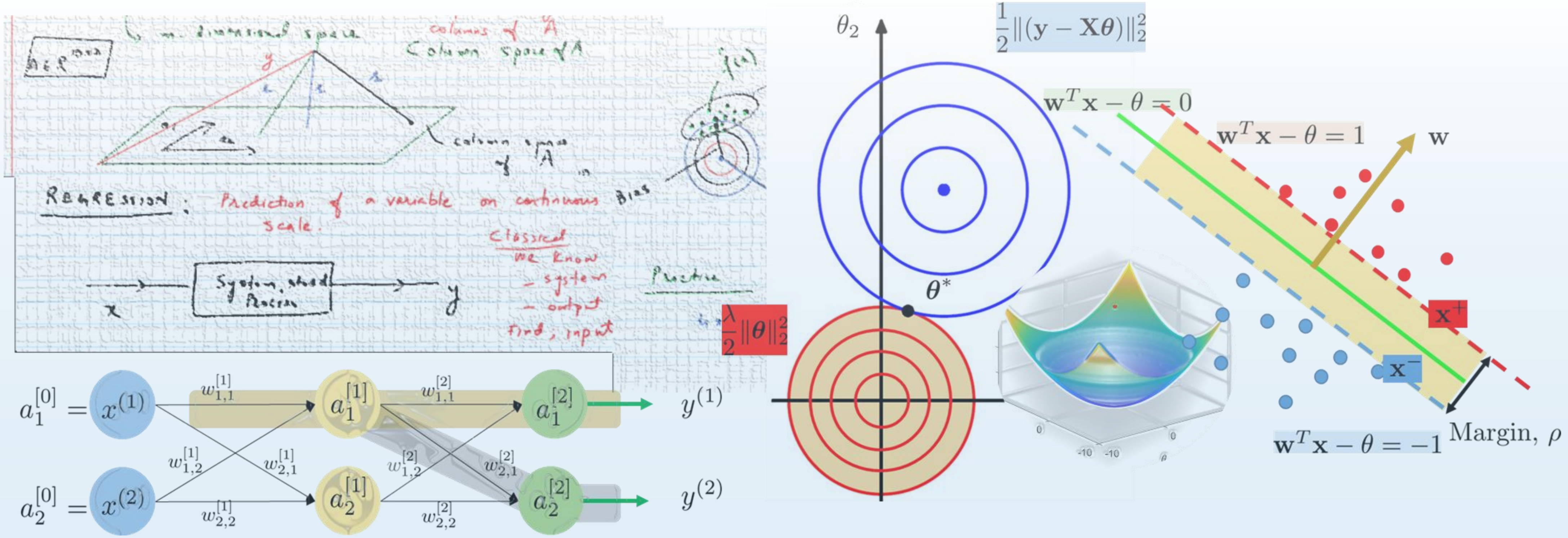


AI-501 Mathematics for AI

Vectors – Notation and Basic Operations

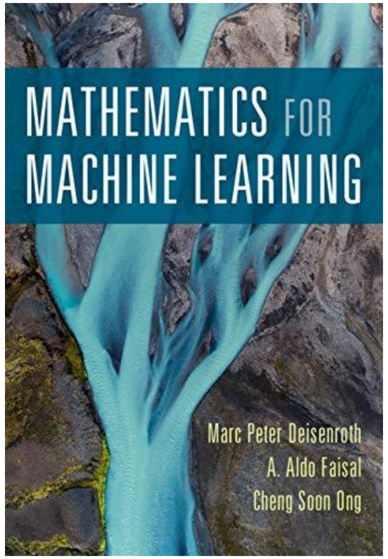
Zubair Khalid
School of Science and Engineering

https://www.zubairkhalid.org/ai501_2024.html

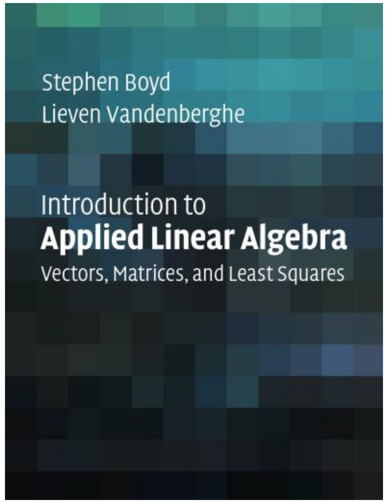
LUMS
A Not-for-Profit University

Outline

- Vectors notation
- Applications
- Basic vector operations



Chapter 2



Chapter 1 and 2

Vectors

Definition:

A vector is an **ordered** finite list of numbers (real or complex).

Notation:

Usually denoted by a letter symbol; stack the list of numbers in an ordered form.

For example, consider a vector of 4 real numbers given by

$$a = \begin{bmatrix} -1.1 \\ 30.1 \\ 6.1 \\ -2.7 \end{bmatrix} \qquad a = \begin{pmatrix} -1.1 \\ 30.1 \\ 6.1 \\ -2.7 \end{pmatrix} \qquad a = (-1.1,\, 30.1,\, 6.1,\, -2.7)$$

Size of a vector: Number of elements the vector contains (also referred to as length or dimension).

We usually express vector b of size n as $b \in \mathbf{R}^n$ and call it n-vector.

Entry of a vector: b_k - k-th entry of the vector b. For example, $a_2 = 30.1$ for the vector a defined above.

Vectors

Zero vector: A vector with all elements equal to zero.
denoted by $\mathbf{0} \in \mathbf{R}^n$.

One vector: A vector with all elements equal to one.
denoted by $\mathbf{1} \in \mathbf{R}^n$.

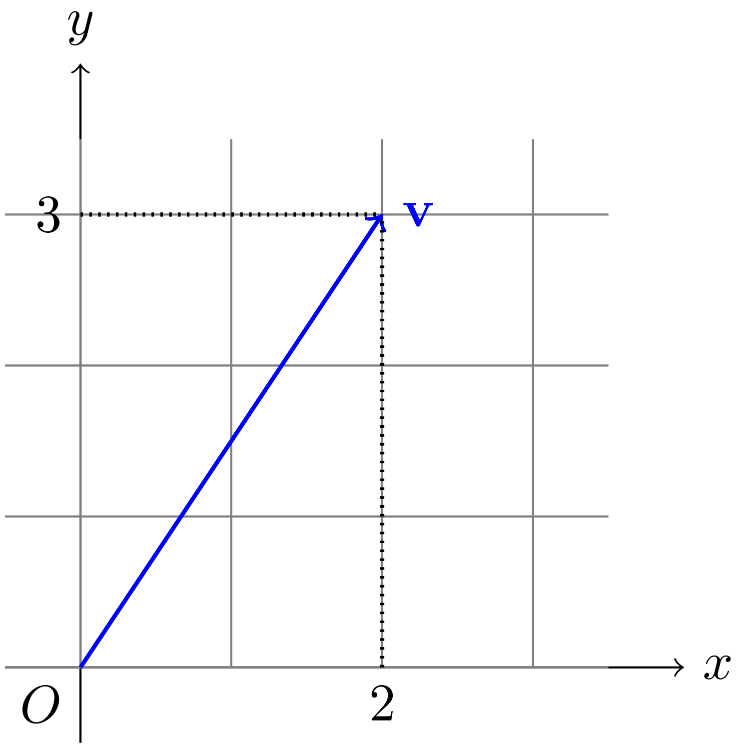
Unit vector: A standard unit vector is vector with all elements zero except one element that is equal to one.
denoted by $e_i \in \mathbf{R}^n$ and is defined as

$$(e_i)_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad = \delta_{i,j}$$

k-Sparse vector: A vector with at-most **k** non-zero entries.

Geometric Interpretation

2-vector: The following diagram illustrates a vector $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ in the $\mathbb{R}^2$ plane:

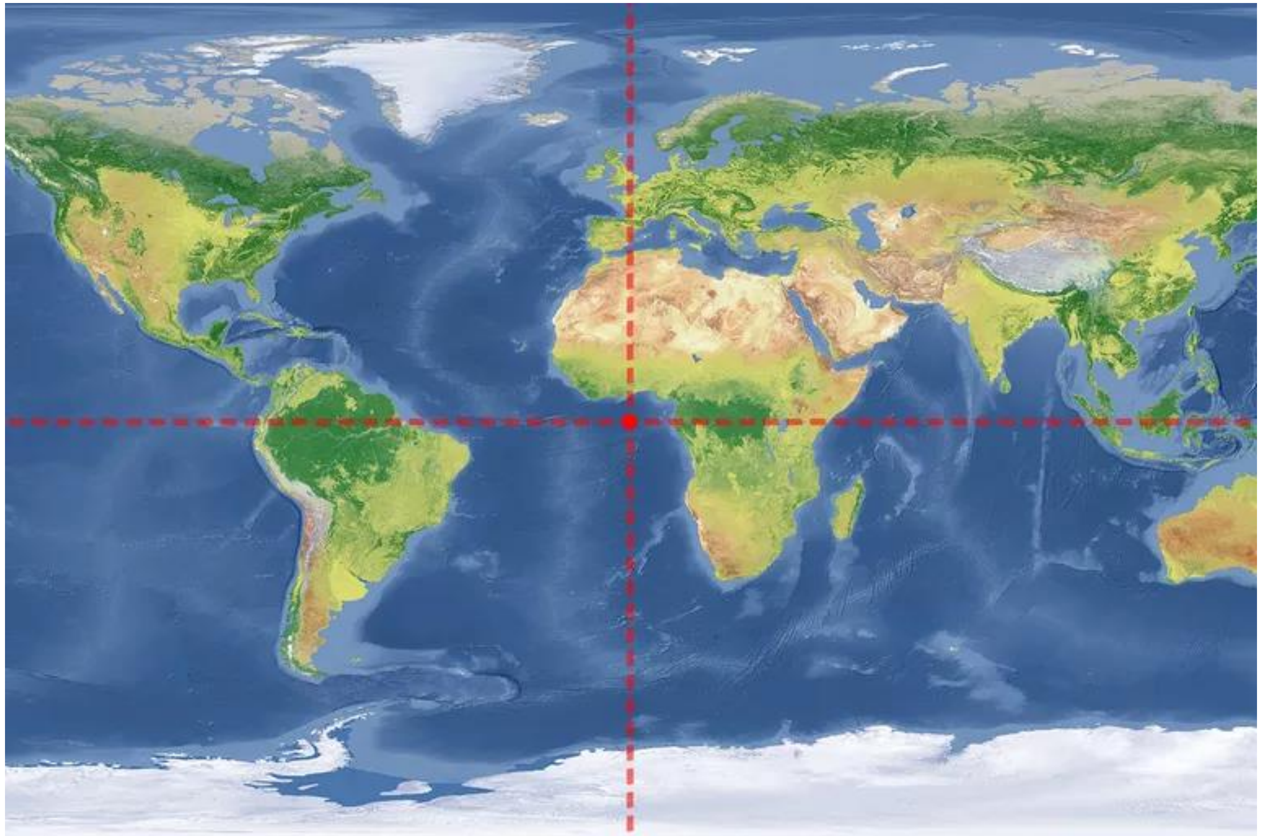


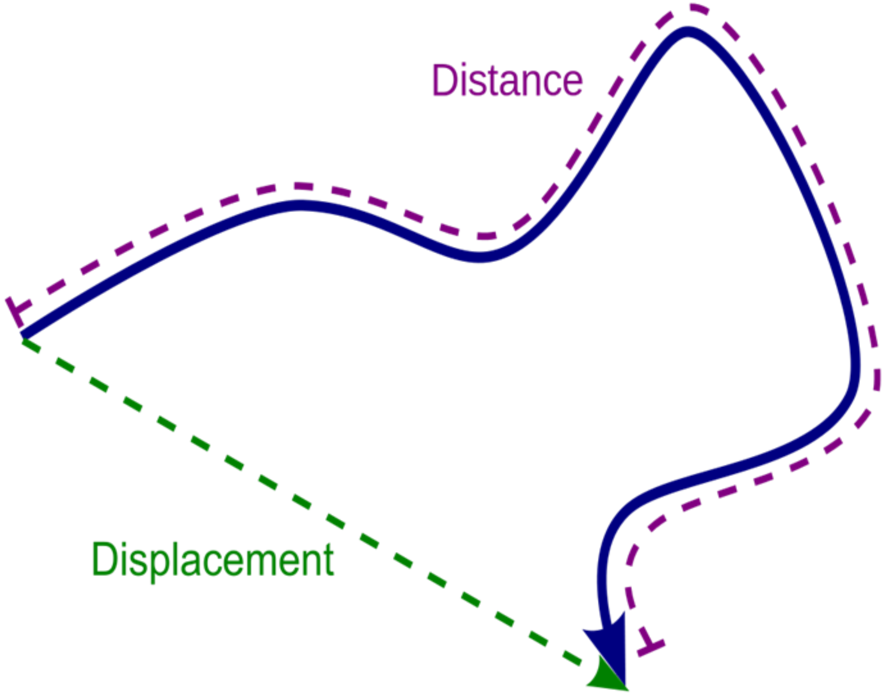
Here, the vector $\mathbf{v}$ is represented geometrically by an arrow from the origin O to the point $(2, 3)$ in the plane.

Examples of Vectors - Applications

Location

Displacement, Velocity, Acceleration





Examples of Vectors - Applications

Color

Each color is represented by 3-vector.

Quantities

An n-vector q can represent the amounts or quantities of n different resources or products held (or produced, or required) by an entity such as a company.

For example, n-vector represents the quantity of n products stocked in a warehouse.

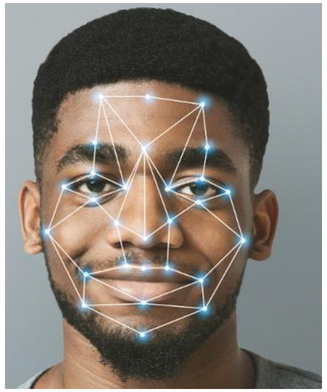
Values across a population

An n-vector can give the values of some quantity across a population of individuals or entities.

For example, an n-vector **a** can represent the blood pressure of a collection of n patients, with a_i the blood pressure of patient i, for i = 1, 2, …, n.

Examples of Vectors - Applications

Image

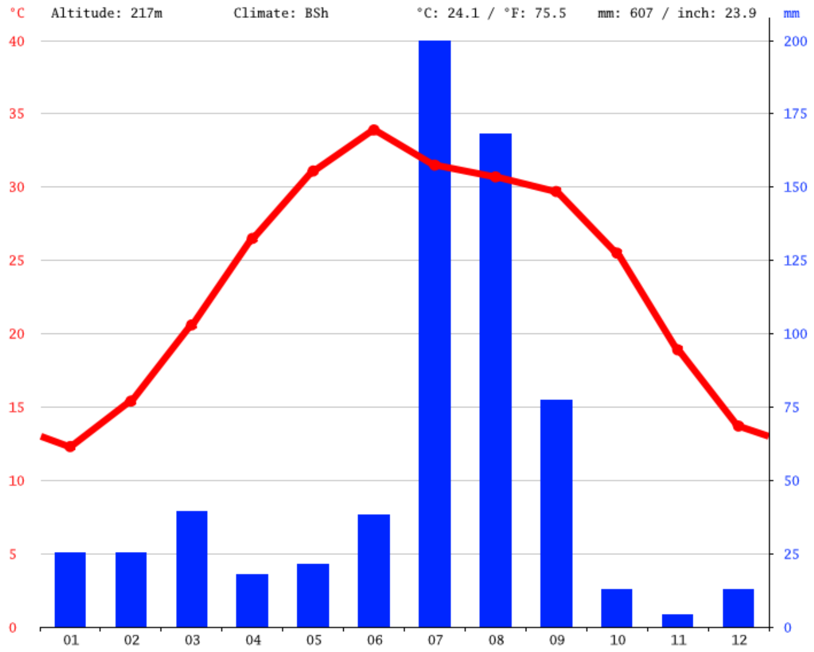


Time series

12-vector can represent the average monthly temperature, rainfall, pressure etc of Lahore.

30-vector can represent the number of expected COVID-19 patients in Pakistan cases over the next 30 days.

Other examples include exchange rate, audio, and, in fact, any quantity that varies over time.

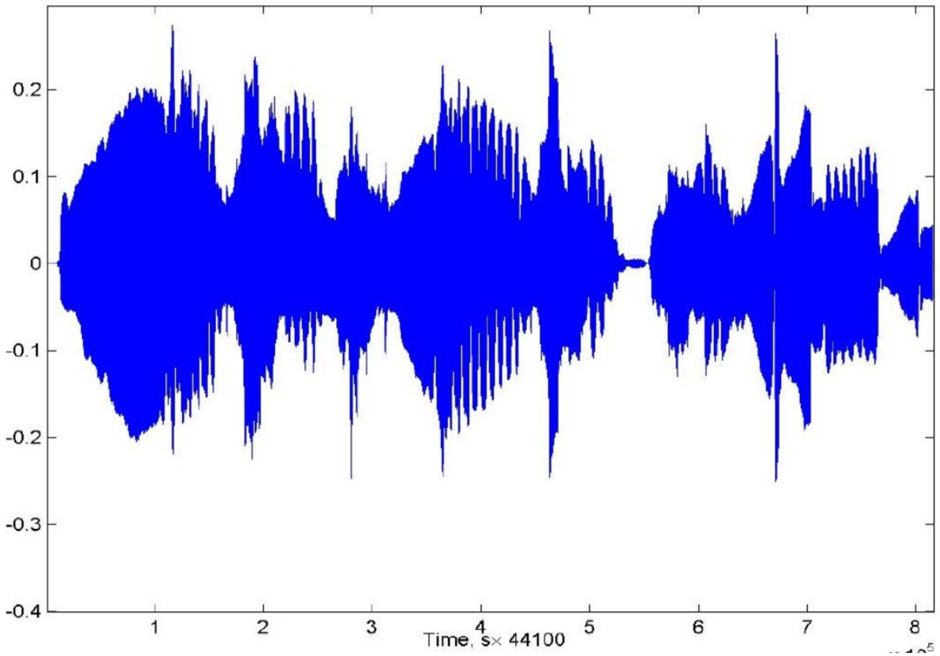


Examples of Vectors - Applications

Audio

Sound (Flute) versus time

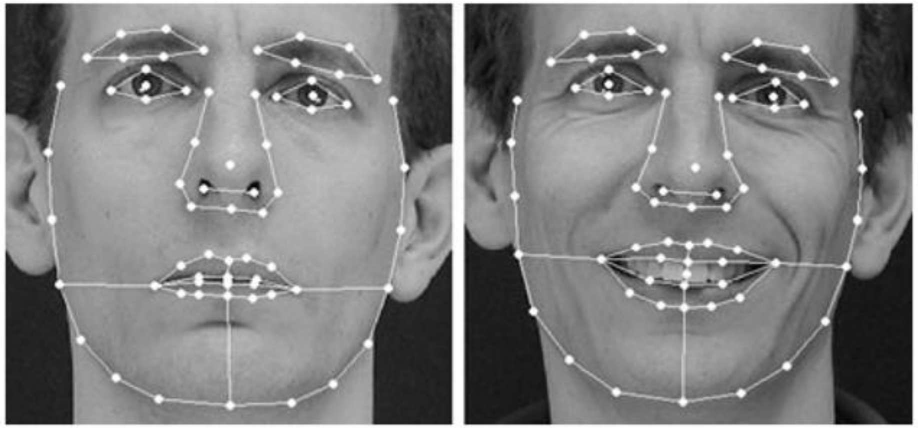
8 second sound = 44100x8-vector



Feature or Attribute

In Machine Learning, classification is mostly carried out collecting features or characteristics derived from the object.

Such a vector is sometimes called a feature vector, and its entries are called the features or attributes.



Emotion Recognition

Operations on Vectors

Additivity and Scaling

Additivity of Vectors: The sum of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is obtained by adding their corresponding components:

$$(\mathbf{u} + \mathbf{v})_i = u_i + v_i, \quad \text{for } i = 1, 2, \ldots, n.$$

Scaling of a Vector by a Scalar: A vector $\mathbf{v} \in \mathbb{R}^n$ can be scaled by a scalar $\alpha \in \mathbb{R}$ by multiplying each component of the vector by the scalar:

$$(\alpha \mathbf{v})_i = \alpha v_i, \quad \text{for } i = 1, 2, \ldots, n.$$

These operations are fundamental in vector algebra, allowing for the combination and modification of vectors in $\mathbb{R}^n$.

Operations on Vectors

Linear Combination

A **linear combination** of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ is an expression formed by multiplying each vector by a scalar and then adding the results. Specifically, a linear combination of these vectors is given by:

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k,$$

where $c_1, c_2, \ldots, c_k \in \mathbb{R}$ are scalars called the coefficients of the linear combination.

The concept of a linear combination is fundamental in linear algebra, as it is used to describe vector spaces, spans, and the solutions to systems of linear equations.

Operations on Vectors

Linear Combination

Affine Combination: An *affine combination* of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ is a linear combination where the sum of the coefficients is 1:

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k,$$

where $c_1, c_2, \ldots, c_k \in \mathbb{R}$ and

$$\sum_{i=1}^{k} c_i = 1.$$

Affine combinations generalize the concept of a linear combination by imposing the constraint on the sum of the coefficients, which preserves the "affine structure" of the space (i.e., translations and weighted averages).

Operations on Vectors

Linear Combination

Convex Combination: A *convex combination* is a special case of an affine combination where all coefficients are non-negative:

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k,$$

where $c_1, c_2, \ldots, c_k \geq 0$ and

$$\sum_{i=1}^{k} c_i = 1.$$

Convex combinations are used to describe points that lie within the "convex hull" of a set of vectors, representing weighted averages or mixtures of the vectors where no weight is negative.

Operations on Vectors

Inner Product

The **inner product** (or dot product) of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is a scalar quantity that measures their similarity. It is defined as:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\top \mathbf{v} = \sum_{i=1}^{n} u_i v_i.$$

Geometrically, the inner product is related to the angle between the vectors and can be used to determine orthogonality, with $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ indicating that $\mathbf{u}$ and $\mathbf{v}$ are perpendicular.

The inner product is fundamental in many aspects of machine learning and AI, serving as a basis for understanding distances, angles, and similarity between vectors in various contexts.

Operations on Vectors

Inner Product - Applications

Similarity and Distance Measurement:
The inner product is commonly used to measure the similarity between vectors. For example, in text analysis, the cosine similarity between two vectors $\mathbf{u}$ and $\mathbf{v}$ is given by:

$$\text{cosine similarity} = \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

A higher inner product indicates greater similarity, which is useful in clustering, classification, and recommendation systems.

Linear Regression and Least Squares:
In linear regression, predictions are computed as the inner product between a vector of features $\mathbf{x}$ and a vector of weights $\mathbf{w}$:

$$\hat{y} = \mathbf{w}^\top \mathbf{x}.$$

Operations on Vectors

Inner Product - Applications

Neural Networks and Backpropagation:
In neural networks, the inner product is used in the forward pass to compute the weighted sum of inputs to neurons:

$$z = \mathbf{w}^\top \mathbf{x} + b.$$

This forms the basis of linear transformations in fully connected layers.

Matrix Factorization for Recommender Systems:
In recommender systems, the inner product is used to predict user-item ratings. Given user vector $\mathbf{u}$ and item vector $\mathbf{v}$, the predicted rating is:

$$\hat{r} = \mathbf{u}^\top \mathbf{v}.$$

Operations on Vectors

Inner Product – Applications – In class example:

Average age in a population. Suppose the 100-vector x represents the distribution of ages in some population of people, with x_i being the number of $i-1$ year olds, for $i = 1, \dots, 100$. (You can assume that $x \neq 0$, and that there is no one in the population over age 99.) Find expressions, using vector notation, for the following quantities.

(a) The total number of people in the population.

(b) The total number of people in the population age 65 and over.

(c) The average age of the population. (You can use ordinary division of numbers in your expression.)

Operations on Vectors

Norm

The **norm** of a vector $\mathbf{v} \in \mathbb{R}^n$ is a measure of its magnitude or length. A norm is a function $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ that satisfies the following properties:

Non-negativity: $\|\mathbf{v}\| \geq 0$ for all $\mathbf{v} \in \mathbb{R}^n$.

Definiteness: $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

Homogeneity (Scaling): $\|\alpha \mathbf{v}\| = |\alpha| \|\mathbf{v}\|$ for any scalar $\alpha \in \mathbb{R}$.

Triangle Inequality: $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

The p-norm (or *Lp norm*) of a vector $\mathbf{v} \in \mathbb{R}^n$ is defined as:

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p}, \quad \text{for } p \geq 1.$$

Operations on Vectors

Norm

The p-norm has special names for different values of p:

1-norm (Manhattan Norm or Taxicab Norm):

$$\|\mathbf{v}\|_1 = \sum_{i=1}^{n} |v_i|.$$

It measures the sum of the absolute values of the components of $\mathbf{v}$.

2-norm (Euclidean Norm):

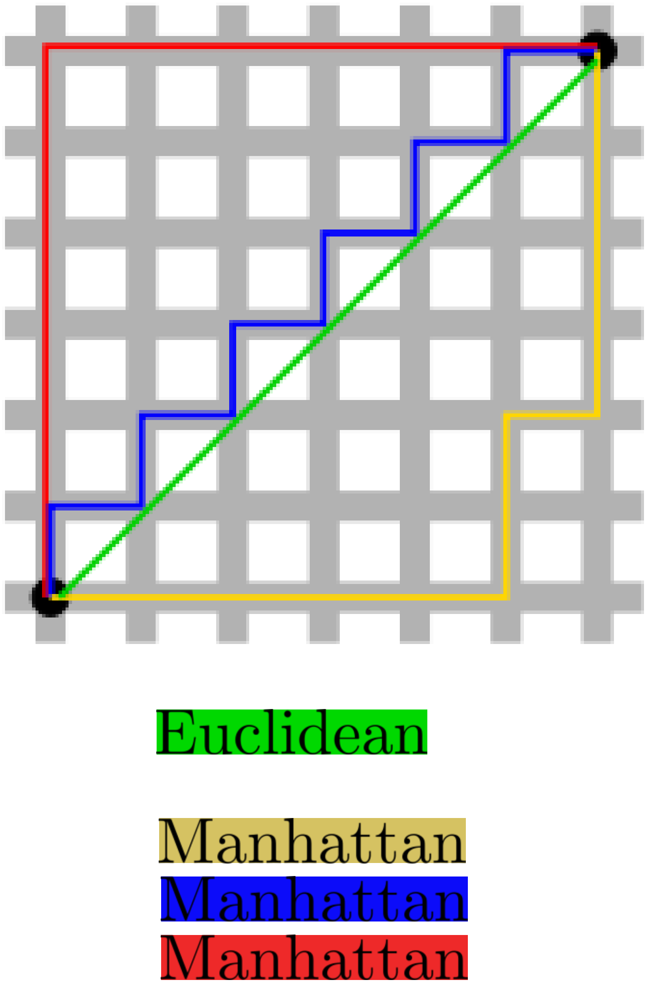
$$\|\mathbf{v}\|_2 = \left(\sum_{i=1}^{n} v_i^2\right)^{1/2}.$$

This is the standard Euclidean length or magnitude of $\mathbf{v}$.

∞-norm (Maximum Norm or Chebyshev Norm):

$$\|\mathbf{v}\|_\infty = \max_{1\leq i\leq n} |v_i|.$$

It measures the maximum absolute value of the components of $\mathbf{v}$.



Operations on Vectors

Norm – Example:

Consider the vector $\mathbf{v} = \begin{bmatrix} 3 & -4 & 2 \end{bmatrix}^\top$. Let's compute different norms of this vector:

1-norm:

$$\|\mathbf{v}\|_1 = |3| + |-4| + |2| = 3 + 4 + 2 = 9.$$

2-norm:

$$\|\mathbf{v}\|_2 = \left(3^2 + (-4)^2 + 2^2\right)^{1/2} = (9 + 16 + 4)^{1/2} = \sqrt{29}.$$

∞-norm:

$$\|\mathbf{v}\|_\infty = \max\left(|3|, |-4|, |2|\right) = 4.$$

Operations on Vectors

Distance

The distance between $\mathbf{u}$ and $\mathbf{v}$ under the p-norm, denoted as $d_p(\mathbf{u}, \mathbf{v})$, is defined as the norm of their difference:

$$d_p(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_p = \left(\sum_{i=1}^{n} |u_i - v_i|^p \right)^{1/p}, \quad \text{for } p \geq 1.$$

Interpretation: *Compare two vectors; close or far*

Operations on Vectors

Application Examples

Prediction error

In regression and other predictive models, the **prediction error** is measured as the distance between the true values $\mathbf{y} \in \mathbb{R}^n$ and the predicted values $\hat{\mathbf{y}} \in \mathbb{R}^n$. The most commonly used distance is the Euclidean distance, which gives the root mean square error (RMSE):

$$\text{RMSE} = \frac{1}{\sqrt{n}} \|\mathbf{y} - \hat{\mathbf{y}}\|_2 = \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \right)^{1/2}.$$

The RMSE provides a measure of how closely the predicted values match the true values.

Alternatively, the mean absolute error (MAE) uses the Manhattan distance:

$$\text{MAE} = \frac{1}{n} \|\mathbf{y} - \hat{\mathbf{y}}\|_1 = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|.$$

Operations on Vectors

Angle $a, b \in \mathbf{R}^n$

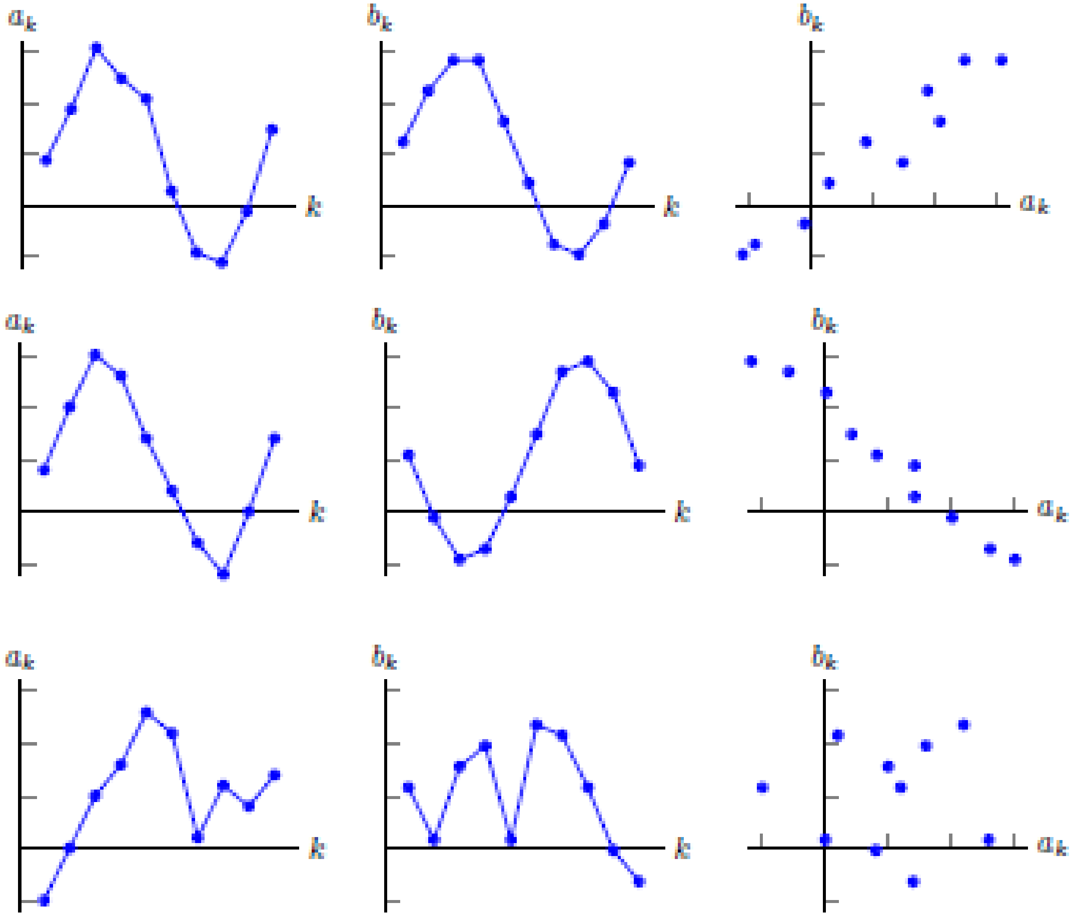
$$\theta = \angle(a, b) = \cos^{-1} \frac{a^T b}{||a|| \, ||b||}$$

Operations on Vectors

Correlation coefficient $a, b \in \mathbf{R}^n$

$$\tilde{a} = a - \mathbf{avg}(a)\mathbf{1}, \qquad \tilde{b} = b - \mathbf{avg}(b)\mathbf{1}$$

$$\rho = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\| \, \|\tilde{b}\|}$$



Operations on Vectors

Standard Deviation

The **standard deviation** of a vector $\mathbf{x} \in \mathbb{R}^n$ is a measure of the spread or dispersion of its elements around their average. It indicates how much the components of the vector vary from the average value.

Let $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^\top$ be a vector of n elements. The average of the vector, denoted as avg, is given by:

$$\text{avg} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

The standard deviation σ of the vector is defined as:

$$\text{std}(x) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \text{avg})^2}.$$

This formula calculates the square root of the average of the squared deviations from the average value, providing a measure of how much the components of the vector deviate from the average.

Operations on Vectors

Standard Deviation

Relationship between std, avg and rms

The relationship between rms, avg, and std is given by:

$$\mathrm{rms}^2 = \mathrm{avg}^2 + \mathrm{std}^2.$$

This shows that the square of the rms is the sum of the square of the average and the variance of the vector.

Operations on Vectors

Standard Deviation

Relationship between std, avg and rms

In machine learning, the performance of a model is often evaluated by its prediction errors, which can be decomposed into three key components: bias, variance, and noise. These components are closely related to the root mean square (RMS) error, which measures the accuracy of the model's predictions.
The **RMS error** of a model is a measure of how much the model's predictions differ from the true values. For a set of predictions $\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 & \hat{y}_2 & \cdots & \hat{y}_n \end{bmatrix}^\top$ and true values $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^\top$, the RMS error is defined as:

$$\text{RMS Error} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}.$$

The RMS error is often used as a loss function to measure the performance of regression models.

Operations on Vectors

Standard Deviation

Relationship between std, avg and rms

The error in a model's predictions can be broken down into **bias** and **variance**:

Bias: Bias refers to the error introduced by approximating a real-world problem with a simplified model. It measures the difference between the expected prediction of the model and the true value:

$$\text{Bias}(\mathbf{x}) = \mathbb{E}[\hat{y}(\mathbf{x})] - y(\mathbf{x}),$$

where $\hat{y}(\mathbf{x})$ is the prediction of the model for input $\mathbf{x}$, and $y(\mathbf{x})$ is the true value.

Variance: Variance captures the model's sensitivity to variations in the training data. It measures the spread of the model's predictions around the expected prediction:

$$\text{Variance}(\mathbf{x}) = \mathbb{E}[(\hat{y}(\mathbf{x}) - \mathbb{E}[\hat{y}(\mathbf{x})])^2].$$

A high variance indicates that the model's predictions vary greatly with different training datasets, leading to overfitting.

Operations on Vectors

Standard Deviation

Relationship between std, avg and rms

The **bias-variance trade-off** is a fundamental concept in machine learning, which states that there is a trade-off between the bias and variance of a model: - A model with **high bias** typically makes simplistic assumptions, leading to underfitting and poor predictive performance. - A model with **high variance** is too sensitive to the training data, leading to overfitting and poor generalization to new data.

The RMS error can be decomposed into the bias and variance components of the model's predictions, along with the inherent noise (irreducible error) in the data. This is known as the bias-variance decomposition:

$$\text{RMS Error}^2 = \text{Bias}^2 + \text{Variance} + \text{Noise}.$$

Here: - Bias^2 is the square of the model's bias. - Variance is the variability of model predictions. - Noise is the irreducible error, representing the variance in the true data distribution that cannot be explained by the model.

Balancing bias and variance is crucial for minimizing the RMS error and achieving good model performance.

Operations on Vectors

Standard Deviation

Concept of Standardization

Let $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^\top$ be a vector of n elements. The standardized vector $\mathbf{z}$ is obtained by subtracting the average (avg) and dividing by the standard deviation (std) of $\mathbf{x}$.

The standardized vector $\mathbf{z}$ is then computed as:

$$\mathbf{z} = \frac{\mathbf{x} - \text{avg}}{\text{std}},$$

where the subtraction and division are element-wise operations.

Properties and Applications

- The standardized vector $\mathbf{z}$ has a mean of 0 and a standard deviation of 1.
- Standardization is commonly used in machine learning and data analysis to improve the performance of algorithms that are sensitive to the scale of features, such as gradient descent optimization, support vector machines, and k-means clustering.

Span

The **span** of a set of vectors is the collection of all possible linear combinations of those vectors. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ are vectors, then the span is the set of all vectors that can be expressed as:

$$\text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k,\ c_i \in \mathbb{R}\}\,.$$

In other words, the span is the subspace of $\mathbb{R}^n$ formed by all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. We will review the subspace concept later.

Properties and Applications

- The span of vectors provides a way to understand the dimensionality and structure of the space they cover. For example, if the vectors are linearly independent, the dimension of the span is equal to the number of vectors.
- The span is used to define vector spaces and subspaces in linear algebra, and plays a key role in solving systems of linear equations, understanding linear transformations, and performing dimensionality reduction in data analysis.

Span

Example

If $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ in $\mathbb{R}^3$, then their span is the set of all vectors in the xy-plane:

$$\text{span}(\mathbf{v}_1, \mathbf{v}_2) = \left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \mid a, b \in \mathbb{R} \right\}.$$

Linear Independence

A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ is said to be **linearly independent** if no vector in the set can be expressed as a linear combination of the others. In other words, the only solution to the equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

is the trivial solution $c_1 = c_2 = \cdots = c_k = 0$, where $\mathbf{0}$ is the zero vector in $\mathbb{R}^n$.

If a set of vectors is not linearly independent, it is said to be **linearly dependent**, meaning at least one vector in the set can be written as a linear combination of the others.

Properties and Applications

- Linear independence is a fundamental concept in linear algebra, as it determines the **basis** of a vector space. A set of vectors that is linearly independent and spans the space forms a basis (to be discussed in detail shortly).
- Determining linear independence is crucial in solving systems of linear equations, finding the rank of a matrix, and understanding the dimensions of vector spaces.

Linear Independence

Example

Consider the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$. These vectors are linearly independent because the only solution to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$$

is $c_1 = c_2 = 0$. Hence, $\{\mathbf{v}_1, \mathbf{v}_2\}$ forms a basis for $\mathbb{R}^2$.

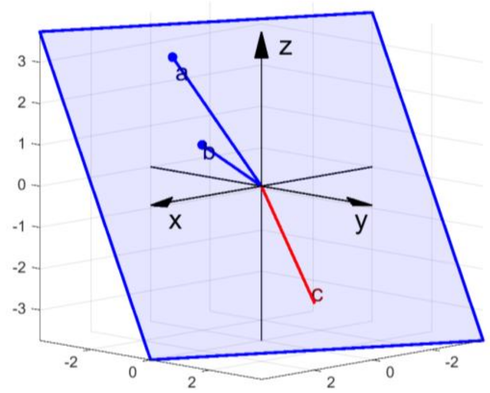
In contrast, if we consider $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, these vectors are linearly dependent because $\mathbf{u}_2 = 2\mathbf{u}_1$.

Linear Independence

Examples

$$a = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \qquad c = \begin{bmatrix} -0.3 \\ 1.5 \\ -2.7 \end{bmatrix}$$

$$1.2a - 0.9b + c = 0$$



Basis

Independence-Dimension Inequality
(Connection with linear independence):

The **independence-dimension inequality** is a fundamental result in linear algebra that relates the number of linearly independent vectors to the dimension of the vector space they inhabit. Specifically, it states that:

> The number of linearly independent vectors in a vector space V cannot exceed the dimension of V.

Mathematically, if $\dim(V) = n$, then any set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subset V$ is linearly independent only if $k \leq n$. If $k > n$, then the set of vectors is guaranteed to be linearly dependent.

Implications

- In practical terms, it provides a way to verify the maximum number of linearly independent vectors that can exist in any subspace. For example, in $\mathbb{R}^3$, any set of more than 3 vectors must be linearly dependent.

Basis

Independence-Dimension Inequality – Example:

(Connection with linear independence):

Consider the vector space $\mathbb{R}^2$, which has dimension 2. According to the independence-dimension inequality, any set of more than 2 vectors in $\mathbb{R}^2$ will be linearly dependent. For instance, the set of vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$ is linearly dependent because the third vector can be written as a combination of the first two:

$$\mathbf{v}_3 = \mathbf{v}_1 + \mathbf{v}_2.$$

Thus, the independence-dimension inequality is a key principle in understanding the structure and limitations of vector spaces.

Basis

Basis of a set of Vectors:

A **basis** is a set of vectors that has two key properties: it is both **linearly independent** and **spans** a certain space. These properties ensure that the set of vectors forms a "building block" from which all other vectors in that space can be uniquely represented.

Formally, a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a basis if:

The vectors are linearly independent, meaning none of the vectors in the set can be expressed as a linear combination of the others.

The vectors span the space, implying any vector in that space can be written as a linear combination of the basis vectors:

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k,$$

where $c_1, c_2, \ldots, c_k \in \mathbb{R}$ are scalars.

- A basis allows every vector in the space to be represented **uniquely** as a linear combination of the basis vectors.

Basis

Basis of a set of Vectors – Example:

Consider the vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$. The set $\{\mathbf{e}_1, \mathbf{e}_2\}$ forms a basis for $\mathbb{R}^2$ because:

The vectors are linearly independent.

Any vector $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$ in $\mathbb{R}^2$ can be written as a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$:

$$\mathbf{v} = a\mathbf{e}_1 + b\mathbf{e}_2.$$

Thus, $\{\mathbf{e}_1, \mathbf{e}_2\}$ is a basis for $\mathbb{R}^2$, and the coefficients a and b uniquely determine any vector in this space.

Orthonormal Vectors

The concepts of orthogonal and orthonormal vectors are fundamental in understanding the geometric relationships between vectors.

Orthogonal Vectors

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are said to be **orthogonal** if their inner (dot) product is zero:

$$\mathbf{u}^\top \mathbf{v} = 0.$$

Geometrically, this means the vectors are at a right angle to each other.

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is called orthogonal if every pair of distinct vectors in the set is orthogonal:

$$\mathbf{v}_i^\top \mathbf{v}_j = 0, \quad \text{for all } i \neq j.$$

Orthonormal Vectors

A set of vectors is called **orthonormal** if it is orthogonal and, in addition, each vector has a norm (length) of 1:

$$\|\mathbf{v}_i\|_2 = 1 \quad \text{for all } i.$$

In other words, the vectors are orthogonal to each other and are unit vectors.

Orthonormal Vectors

Orthogonal and orthonormal vectors are widely used in linear algebra and machine learning. Orthonormal vectors, in particular, simplify many computations, such as projections, transformations, and finding solutions to linear equations, due to their nice properties of orthogonality and unit length.

Example:

Consider the vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$:

These vectors are orthogonal because their inner product is zero:

$$\mathbf{e}_1^\top \mathbf{e}_2 = 1 \cdot 0 + 0 \cdot 1 = 0.$$

They are also orthonormal since each vector has a norm of 1:

$$\|\mathbf{e}_1\|_2 = 1, \quad \|\mathbf{e}_2\|_2 = 1.$$

Orthonormal Vectors

Linear Combination of Orthonormal Vectors

Significance of Orthonormal Vectors

Expressing a vector as a linear combination of orthonormal vectors has significant advantages in both theoretical understanding and practical applications in linear algebra and machine learning.

Simplified Coefficients (Projections):

When a vector $\mathbf{u} \in \mathbb{R}^n$ is expressed as a linear combination of orthonormal vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, the coefficients of this combination are simply the inner products (projections) of $\mathbf{u}$ onto each orthonormal vector:

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k, \quad \text{where } c_i = \mathbf{v}_i^\top \mathbf{u}.$$

This property makes it easy to calculate the coefficients without solving a system of equations, as would be required for general bases.

Uniqueness of Representation:

A vector expressed as a linear combination of orthonormal vectors has a unique set of coefficients. This uniqueness is particularly useful in many applications, such as signal processing, where signals are decomposed into orthonormal basis functions (e.g., Fourier series).

Orthonormal Vectors

Linear Combination of Orthonormal Vectors

Example: Decomposing a Vector in $\mathbb{R}^3$

Consider the vector $\mathbf{u} = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$ and an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

The coefficients of $\mathbf{u}$ in this orthonormal basis are simply:

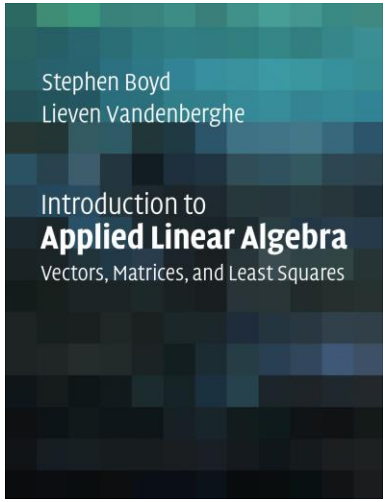
$$c_1 = \mathbf{v}_1^\top \mathbf{u} = 3, \quad c_2 = \mathbf{v}_2^\top \mathbf{u} = 4, \quad c_3 = \mathbf{v}_3^\top \mathbf{u} = 0.$$

Thus, the vector $\mathbf{u}$ can be expressed as:

$$\mathbf{u} = 3\mathbf{v}_1 + 4\mathbf{v}_2 + 0\mathbf{v}_3.$$

Outline

- *Gram-Schmidt Orthogonalization*
  - *Overview*
  - *Algorithm*
  - *Example*



Chapter 5

Gram-Schmidt Orthogonalization

Overview: **Input:** Given a list of vectors $a_1, a_2, \ldots, a_k$

Output:

1. Test linear independence
2. Orthonormal vectors $q_1, q_2, \ldots, q_k$

Gram-Schmidt Orthogonalization

Example: $\mathbf{a}_1 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

$\tilde{q}_1 = a_1$

$$\tilde{q}_1 = \mathbf{a}_1 = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \qquad \|\tilde{q}_1\| = \sqrt{4^2 + 3^2} = \sqrt{16+9} = \sqrt{25} = 5 \qquad q_1 = \frac{1}{5}\begin{pmatrix} 4 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{4}{5} \\ \frac{3}{5} \end{pmatrix}$$

$\tilde{q}_2 = a_2 - (q_1^T a_2)\, q_1$ We now subtract the projection of $\mathbf{a}_2$ onto q_1.

Gram-Schmidt Orthogonalization

Example: $\mathbf{a}_1 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

$\tilde{q}_1 = a_1$

$$\tilde{q}_1 = \mathbf{a}_1 = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \qquad \|\tilde{q}_1\| = \sqrt{4^2 + 3^2} = \sqrt{16 + 9} = \sqrt{25} = 5 \qquad q_1 = \frac{1}{5}\begin{pmatrix} 4 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{4}{5} \\ \frac{3}{5} \end{pmatrix}$$

$\tilde{q}_2 = a_2 - (q_1^T a_2)\, q_1$ We now subtract the projection of $\mathbf{a}_2$ onto q_1.

$$\mathrm{proj}_{q_1}(\mathbf{a}_2) = \left(\frac{\mathbf{a}_2^T q_1}{q_1^T q_1}\right) q_1 = (-1)\begin{pmatrix} \frac{4}{5} \\ \frac{3}{5} \end{pmatrix} = \begin{pmatrix} \frac{-4}{5} \\ \frac{-3}{5} \end{pmatrix}$$

$$\tilde{q}_2 = \mathbf{a}_2 - \mathrm{proj}_{q_1}(\mathbf{a}_2) = \begin{pmatrix} -2 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{-4}{5} \\ \frac{-3}{5} \end{pmatrix} = \begin{pmatrix} \frac{-10}{5} + \frac{4}{5} \\ \frac{5}{5} + \frac{3}{5} \end{pmatrix} = \begin{pmatrix} \frac{-6}{5} \\ \frac{8}{5} \end{pmatrix}$$

$$\|\tilde{q}_2\| = \sqrt{\left(\frac{-6}{5}\right)^2 + \left(\frac{8}{5}\right)^2} = \sqrt{\frac{36}{25} + \frac{64}{25}} = \sqrt{\frac{100}{25}} = \sqrt{4} = 2 \qquad q_2 = \frac{1}{2}\begin{pmatrix} \frac{-6}{5} \\ \frac{8}{5} \end{pmatrix} = \begin{pmatrix} \frac{-3}{5} \\ \frac{4}{5} \end{pmatrix}$$

Gram-Schmidt Orthogonalization

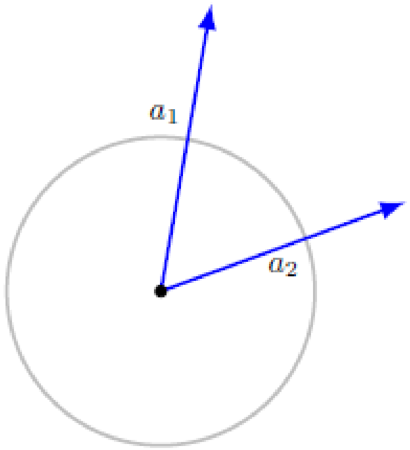
Interpretation

Orthogonalization

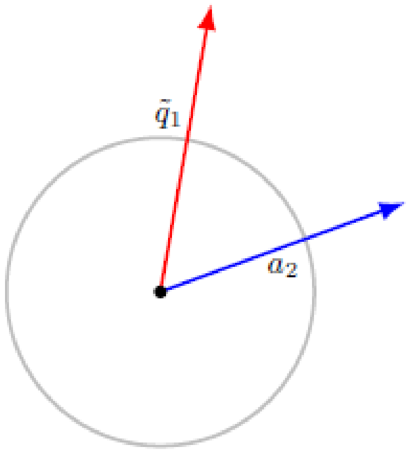
$$\tilde{q}_i = a_i - \sum_{j=1}^{i-1} (q_j^T a_i)\, q_j$$

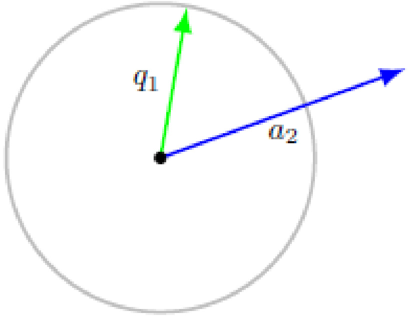
Normalization

$$q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$$



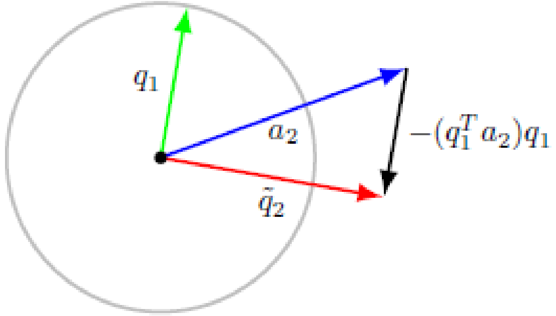
$$\tilde{q}_1 = a_1$$

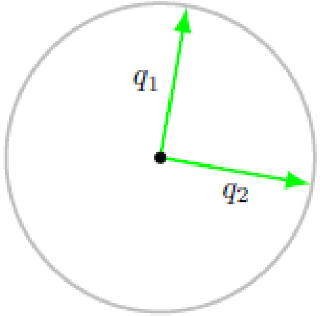




Iteration 1

$$\tilde{q}_2 = a_2 - (q_1^T a_2)\, q_1$$





Iteration 2

Gram-Schmidt Orthogonalization

Algorithm:

Input: Given a list of vectors $a_1, a_2, \ldots, a_k$

Output:

1. Test linear independence
2. Orthonormal vectors $q_1, q_2, \ldots, q_k$

$$\tilde{q}_i = a_i - \sum_{j=1}^{i-1} (q_j^T a_i)\, q_j$$

Orthogonalization

$$q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$$

Normalization

Example for $k = 3$

- i=1 $\quad \tilde{q}_1 = a_1$
- i=2 $\quad \tilde{q}_2 = a_2 - (q_1^T a_2)\, q_1$
- i=3 $\quad \tilde{q}_3 = a_3 - \left((q_1^T a_3)\, q_1 + (q_2^T a_3)\, q_2 \right)$

If $\tilde{q}_i = 0$, we detect linear dependence.

Outline

- Vector Spaces
- Subspaces
- Dimension of subspace

Vector Space

Definition

Two operators:

Let V be a set of vectors.

($+$) Addition of vectors
($\cdot$) multiplication with scalar

For $u, v, w \in V$ and $a, b \in \mathbf{R}$, following axioms hold

1. $u + v \in V$
2. $u + v = v + u$
3. $(u + v) + w = v + (u + w)$
4. $0 \in V : \; v + 0 = v$
5. $\exists -u \in V : \; u + (-u) = 0$
6. $a\,u \in V$
7. $a\,(u + v) = a\,u \; + \; a\,v$
8. $(a + b)\,u = a\,u \; + \; b\,u$
9. $(ab)\,u = a\,(b\,u)$
10. $(1)\,u = u$

Subspace

Definition

- V is a vector space

- W is a subset of V

- If W is also a vector space under the addition and scalar multiplication that is defined on V

then W is a subspace

Alternatively

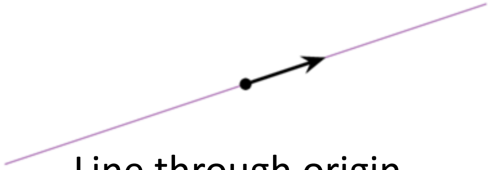
W, a subset of V, is a subspace if

- $0 \in W$ Non-emptiness
- For $u, v \in W$, $u + v \in W$ Closure under addition
- For $a \in R$ and $u \in W$, $au \in W$ Closure under multiplication

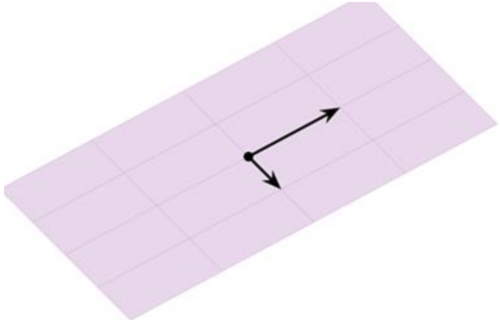
- Every vector space V has at least two subspaces. Namely, V itself and $\{0\}$ (the zero subspace).

Subspace

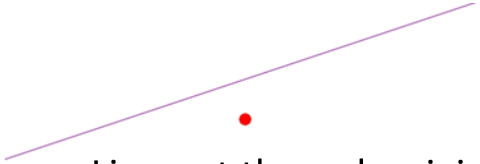
Examples (Subspace or not):



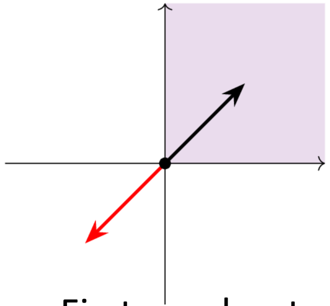
Line through origin



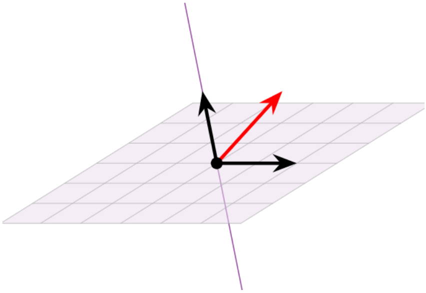
Plane through origin



Line not through origin



First quadrant



Line union plane

Subspace

Examples – In class questions:

(a) Let W be the set of all points, (x, y), from $\mathbb{R}^2$ in which $x \geq 0$. Is this a subspace of $\mathbb{R}^2$?

(b) Let W be the set of all points from $\mathbb{R}^3$ of the form $(0, x_2, x_3)$. Is this a subspace of $\mathbb{R}^3$?

Dimension of Subspace

The number of vectors in any basis of *V* is called the ***dimension*** of *V.*

Expressed as dim(V)

Examples

- line through origin in $\mathbf{R}^4$

- plane through origin in $\mathbf{R}^3$