UNIVERSITY OF MARYLAND

MSML610: Advanced Machine Learning

Linear Algebra

Instructor: Dr. GP Saggese - `gsaggese@umd.edu`

References:

Linear algebra

- **Linear algebra**
  - Vector and vector spaces
  - Affine spaces
  - Linear functions
  - Linear dependence
  - Basis
  - Dimension of a vector space
  - Direct sum
  - Connections between Machine Learning and Linear Algebra

Vector and vector spaces

- Linear algebra
  - **Vector and vector spaces**
  - Affine spaces
  - Linear functions
  - Linear dependence
  - Basis
  - Dimension of a vector space
  - Direct sum
  - Connections between Machine Learning and Linear Algebra

Field: definition

A field $\mathbb{F} = (X, +, *)$ is a set X with two binary operations $+$ and *, satisfying the following 6 axioms:

1. Closed with respect to $+$ and *:

$$a, b \in X \implies a + b \in X$$
$$a, b \in X \implies a * b \in X$$

2. Commutativity of $+$ and *:

$$a + b = b + a$$
$$a * b = b * a$$

3. Associativity of $+$ and *:

$$a + (b + c) = (a + b) + c = a + b + c$$
$$a * (b * c) = (a * b) * c = a * b * c$$

4. Distributivity of multiplication over addition:

$$a * (b + c) = a * b + a * c$$

5. Existence of $+$ and * identity elements, 0 and 1:

$$a + 0 = a$$

Field: examples

- Examples
  - The set of $\mathbb{R}, \mathbb{C}$, GF(2)
  - The set of rational numbers $\mathbb{Q}$, i.e., numbers that can be written as fraction $\frac{a}{b}$ with $a, b \in \mathbb{Z}$ and $b \neq 0$
- Non-examples
  - The set of positive integers $\mathbb{N} = 1, 2, 3, \ldots$ is not a field
  - The set of integers $\mathbb{Z} = \ldots$, -2, -1, 0, 1, 2, $\ldots$\$ is not a field

Vector space: definition

- A "vector space $\mathcal{V}$ over a field $\mathbb{F}$" is a triple $(\mathcal{V}, \mathbb{F}, +, \cdot)$ where:
  - $\mathcal{V}$ is a set of vectors
  - $\mathbb{F}$ is a field of scalars
  - $+$ is a sum operation between vectors
  - $\cdot$ is a scalar multiplication
- A vector space needs to satisfy the following 2 properties:
  1. Closed with respect to to scalar multiplication: if $\underline{\boldsymbol{x}} \in \mathcal{V}$, then $\alpha \cdot \underline{\boldsymbol{x}} \in \mathcal{V}$
  2. Closed with respect to to vector addition: if $\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \in \mathcal{V}$, then $\underline{\boldsymbol{x}} + \underline{\boldsymbol{y}} \in \mathcal{V}$

Linear combination of vectors

- The vector

$$\alpha_1 \underline{\boldsymbol{v}}_1 + \alpha_2 \underline{\boldsymbol{v}}_2 + ... + \alpha_n \underline{\boldsymbol{v}}_n$$

  is a linear combination of vectors $\underline{\boldsymbol{v}}_1, ..., \underline{\boldsymbol{v}}_n$ with coefficients $\alpha_1, ..., \alpha_n$

- A linear combination can be written in matrix form:

$$\underline{\underline{\boldsymbol{V}}} \cdot \underline{\boldsymbol{\alpha}} \text{ or } \underline{\boldsymbol{\alpha}}^T \cdot \underline{\underline{\boldsymbol{V}}}^T$$

  where $\underline{\underline{\boldsymbol{V}}} = (\underline{\boldsymbol{v}}_1 | ... | \underline{\boldsymbol{v}}_n)$

Span of vectors

- The span of n m-dimensional vectors is the set of all linear combinations of the n vectors:

$$\text{Span}(\underline{\boldsymbol{v}}_1, ..., \underline{\boldsymbol{v}}_n) = \{\underline{\underline{\boldsymbol{V}}} \cdot \underline{\boldsymbol{\alpha}} \text{ with } \underline{\boldsymbol{\alpha}} \in \mathbb{F}^n\} = \{\underline{\boldsymbol{v}} \in \mathbb{F}^m : \underline{\boldsymbol{v}} = \sum_{i=1}^{n} \alpha_i \underline{\boldsymbol{v}}_i\}$$

- E.g., the span of vectors is a vector space

Null space of a matrix

- Null space of the columns of a matrix $\underline{\underline{\boldsymbol{A}}}$ is defined as the set:

$$\mathrm{Null}(\underline{\underline{\boldsymbol{A}}}) = \{\underline{\boldsymbol{v}} : \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{v}} = \underline{\mathbf{0}}\}$$

  - In words, all the vectors that are coefficients of linear combinations of columns of $\underline{\underline{\boldsymbol{A}}}$ yielding the zero vector
- Null space is a vector space

Homogeneous linear system associated with null space

- From the definition of matrix-vector multiplication the vector $\underline{\boldsymbol{v}}$ is in $\mathrm{Null}(\underline{\underline{\boldsymbol{A}}}) \iff \underline{\boldsymbol{v}}$ is a solution of the homogeneous linear system involving the columns of $\underline{\underline{\boldsymbol{A}}}$:

$$
\begin{aligned}
\underline{\boldsymbol{a}}_1^T \cdot \underline{\boldsymbol{v}} &= 0 \\
\underline{\boldsymbol{a}}_2^T \cdot \underline{\boldsymbol{v}} &= 0 \\
\ldots& \\
\underline{\boldsymbol{a}}_m^T \cdot \underline{\boldsymbol{v}} &= 0
\end{aligned}
$$

- Note that the notation is a bit confusing since we mean the transpose of the columns $\underline{\boldsymbol{a}}_i$ of $\underline{\underline{\boldsymbol{A}}}$ and not the rows of $\underline{\underline{\boldsymbol{A}}}$

Dot product on a vector space: definition

- Given a field of scalars $\mathbb{F}$ and a vector space $\mathcal{V}$ over $\mathbb{F}$, an inner product is a mapping:
$$\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{F}$$
that satisfies the 3 axioms:
  1. Conjugate symmetry (almost commutativity):
  $$\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \rangle = \overline{\langle \underline{\boldsymbol{y}}, \underline{\boldsymbol{x}} \rangle}$$
  2. Linearity in the first argument:
  $$\langle a\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \rangle = a\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \rangle$$
  $$\langle \underline{\boldsymbol{x}} + \underline{\boldsymbol{y}}, \underline{\boldsymbol{z}} \rangle = \langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{z}} \rangle + \langle \underline{\boldsymbol{y}}, \underline{\boldsymbol{z}} \rangle$$
  3. Positive definitiveness:
  $$\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{x}} \rangle \geq 0$$
  $$\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{x}} \rangle = 0 \implies \underline{\boldsymbol{x}} = \underline{\mathbf{0}}$$

Vector inner product

- Aka "dot product", "scalar product"
- Given $\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \in \mathbb{F}^n$ (i.e., same number of components and also same "label" for each element), the inner product of $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ is defined as:

$$\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \rangle = \underline{\boldsymbol{x}}^T \cdot \underline{\boldsymbol{y}} = \sum_{i=1}^{n} x_i y_i \in \mathbb{F}$$

- $\underline{\boldsymbol{x}}^T \cdot \underline{\boldsymbol{y}}$ is read "x dotted y" or "x transposed y"

Affine spaces

Affine space: definition

- If $\underline{\boldsymbol{c}}$ is a vector and $\mathcal{V}$ is a vector space then

$$\mathcal{A} = \underline{\boldsymbol{c}} + \mathcal{V} = \{\underline{\boldsymbol{c}} + \underline{\boldsymbol{v}} : \underline{\boldsymbol{v}} \in \mathcal{V}\}$$

  is called an affine space
- An affine space is a vector space translated by a point represented by a vector
  - E.g., a plane or a line that do not contain the origin

Affine space: example of plane passing through 3 points

- Given 3 not collinear vectors: $\underline{\boldsymbol{u}}_1$, $\underline{\boldsymbol{u}}_2$, and $\underline{\boldsymbol{u}}_3$, the plane containing the endpoints of the 3 vectors can be represented as $\mathcal{A} = \underline{\boldsymbol{u}}_1 + \mathcal{V}$ where $\mathcal{V} = \mathrm{Span}(\underline{\boldsymbol{u}}_2 - \underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_3 - \underline{\boldsymbol{u}}_1)$
- **Note**: the span of the 3 vectors has dimension 3, but an affine space with dimension 2

Affine combination: definition

- An affine combination is a linear combination of vectors where the sum of the (positive or negative) coefficients is 1:

$$\alpha_1 \underline{\boldsymbol{u}}_1 + \ldots + \alpha_n \underline{\boldsymbol{u}}_n \text{ where } \sum_i \alpha_i = 1$$

- In matrix form: $\underline{\underline{\boldsymbol{U}}} \cdot \underline{\boldsymbol{\alpha}}$ where $\underline{\mathbf{1}}^T \underline{\boldsymbol{\alpha}} = 1$

Affine hull of vectors

- Given vectors $\underline{\boldsymbol{u}}_1, ..., \underline{\boldsymbol{u}}_n$, the set of all affine combinations is the affine hull:

$$\mathcal{A} = \{\underline{\boldsymbol{v}} = \sum_i^n \alpha_i \underline{\boldsymbol{u}}_i : \sum_i \alpha_i = 1\} = \{\underline{\boldsymbol{v}} = \underline{\underline{\boldsymbol{U}}}\underline{\boldsymbol{\alpha}} : \underline{\mathbf{1}}^T \underline{\boldsymbol{\alpha}} = 1\}$$

- The affine hull includes each point because if $\underline{\boldsymbol{\alpha}}$ has a single 1 in position i and all others 0, we get $\underline{\boldsymbol{u}}_i$

Affine hull of vectors is an affine space

- We can write the affine hull of $\underline{\boldsymbol{u}}_1, ..., \underline{\boldsymbol{u}}_n$ as an affine space:

$$\underline{\boldsymbol{u}}_i + \mathrm{Span}(\underline{\boldsymbol{u}}_1 - \underline{\boldsymbol{u}}_i, ..., \underline{\boldsymbol{u}}_n - \underline{\boldsymbol{u}}_i)$$

  - Thus an affine space is an affine hull and vice versa
- This is the dual of "the span of vectors is a vector space"

The solution set of non-homogeneous linear system is empty or affine space

- Consider the solution of a system of non-homogeneous linear equations
$$\{\underline{\boldsymbol{x}} : \underline{\boldsymbol{a}}_1^T \underline{\boldsymbol{x}} = \beta_1, ..., \underline{\boldsymbol{a}}_m^T \underline{\boldsymbol{x}} = \beta_m\} \text{ or in matrix form } \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}} = \underline{\boldsymbol{\beta}}$$
- The solution set is either empty or an affine space
- There are two cases: either the non-homogeneous system has solution, or not
- Consider the case where the system of equations has no solutions (e.g., it is contradictory, e.g., $x = 1, x = 2$), then the solution set is empty
- Consider the case where there is a solution
- Each linear system $\underline{\underline{\boldsymbol{A}}}\underline{\boldsymbol{x}} = \underline{\boldsymbol{\beta}}$ has an associated homogeneous linear system $\underline{\underline{\boldsymbol{A}}}\underline{\boldsymbol{x}} = \underline{\boldsymbol{0}}$
- If $\underline{\boldsymbol{u}}_1$ is a solution of the non-homogeneous system (i.e., $\underline{\underline{\boldsymbol{A}}}\underline{\boldsymbol{u}}_1 = \underline{\boldsymbol{\beta}}$), then any other solution $\underline{\boldsymbol{u}}_2$ is a solution (i.e., $\underline{\underline{\boldsymbol{A}}}\underline{\boldsymbol{u}}_2 = \underline{\boldsymbol{\beta}}$) $\iff$ $\underline{\boldsymbol{u}}_2 - \underline{\boldsymbol{u}}_1$ is in the vector space which is the solution of the homogeneous linear system (i.e., $\underline{\underline{\boldsymbol{A}}}(\underline{\boldsymbol{u}}_1 - \underline{\boldsymbol{u}}_2) = \underline{\boldsymbol{0}}$)

Vector space vs affine space: summary

- Linear combination vs affine combination
- Span of vectors vs affine hull of vectors
  - Span (affine hull) is the set of all linear (affine) combinations
- Vector space vs affine space

Matrix

- A matrix $\underline{\underline{\boldsymbol{A}}} \in \mathbb{F}^{m \times n}$ is a two dimensional array with dimensions $m \times n$ of elements from a field $\mathbb{F}$
- Matrix notation
  - $\underline{\underline{\boldsymbol{A}}} \in \mathbb{R}^{m \times n}$ has m rows and n columns
  - A_{ij} is the element on i-th row and j-th column

$$\underline{\underline{\boldsymbol{A}}} = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \dots & A_{i,j} & \dots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix}$$

  - The convention is first rows and then columns (i.e., y-x instead of the more usual x-y) for both elements and dimensions of a matrix
- Notation for rows and columns in a matrix
  - j-th column is $\underline{\boldsymbol{a}}_j$ or $\underline{\boldsymbol{a}}_{:,j}$ (using `numpy` notation)
  - i-th row is $\underline{\boldsymbol{a}}_i^T$ or $\underline{\boldsymbol{a}}_{i,:}$
  - Fixing a coordinate (e.g., row) one gets the orthogonal indices (e.g., column)

Linear functions

- Linear algebra
  - Vector and vector spaces
  - Affine spaces
  - **Linear functions**
  - Linear dependence
  - Basis
  - Dimension of a vector space
  - Direct sum
  - Connections between Machine Learning and Linear Algebra

Linear functions over vector spaces: definition

- Consider two vector spaces $\mathcal{V}$ and $\mathcal{W}$ over the same field $\mathbb{F}$
- A linear function $f : \mathcal{V} \to \mathcal{W}$ satisfies two properties:
    1. $f(\alpha \underline{\boldsymbol{v}}) = \alpha f(\underline{\boldsymbol{v}})$
    2. $f(\underline{\boldsymbol{u}} + \underline{\boldsymbol{v}}) = f(\underline{\boldsymbol{u}}) + f(\underline{\boldsymbol{v}})$
- Linear functions "push linear combination through":

$$f(\alpha_1 \underline{\boldsymbol{v}}_1 + \ldots + \alpha_n \underline{\boldsymbol{v}}_n) = \alpha_1 f(\underline{\boldsymbol{v}}_1) + \ldots + \alpha_n f(\underline{\boldsymbol{v}}_n)$$

    - Equivalent to the 2 properties of linear functions

Matrix and linear function

- From matrix to linear function
  - Given a matrix $\underline{\underline{\boldsymbol{A}}} \in \mathbb{F}^{n \times m}$ we can define the function:

$$f(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}}$$

  - The function $f()$ maps m-vectors into n-vectors
    - The domain is $\mathbb{F}^m$ for the matrix-vector product to be defined
    - The co-domain is $\mathbb{F}^n$
  - $f(\underline{\boldsymbol{x}})$ is a linear function because of the properties of matrix-vector product
- From linear function to matrix
  - Consider a linear function $f : \mathbb{F}^m \to \mathbb{F}^n$
  - We want to find a matrix $\underline{\underline{\boldsymbol{A}}}$ such that $f(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}}$
  - ***Solution***
    - We know that $\underline{\underline{\boldsymbol{A}}} \in \mathbb{F}^{n \times m}$ from matrix-vector product definition
    - If we compute $\underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{e_i}}$, where $e_i = (0, ..., 0, 1, ...0)$ is the i-th standard generator, we obtain $\underline{\boldsymbol{a}}_i$ (i.e., the i-th column of $\underline{\underline{\boldsymbol{A}}}$)
    - Thus $\underline{\underline{\boldsymbol{A}}}$ is the matrix with columns equal to the standard generators transformed by $f()$

Linear functions: examples and non-examples

- **Identity function** is linear
  - Corresponds to the identity matrix
- **Rotation** is linear transformations
  - Corresponds to an orthonormal matrix
- **Scaling** each coordinate independently is linear transformation
  - Corresponds to a diagonal matrix
- **Translation** is *not* a linear function
  - Since it does not satisfy either of the two linearity properties
    1. $f(\alpha \underline{\boldsymbol{v}}) = \alpha f(\underline{\boldsymbol{v}})$
    2. $f(\underline{\boldsymbol{u}} + \underline{\boldsymbol{v}}) = f(\underline{\boldsymbol{u}}) + f(\underline{\boldsymbol{v}})$

Kernel of a Linear Function

- Any linear function $f : \mathcal{V} \to \mathcal{W}$ maps the zero vector of $\mathcal{V}$ to the zero vector of $\mathcal{W}$
- The **kernel** of a linear function f is the set of vectors that are transformed by f into the $\underline{\mathbf{0}}$ vector

$$\mathrm{Ker}(f) = \{\underline{\boldsymbol{v}} : f(\underline{\boldsymbol{v}}) = \underline{\mathbf{0}}\}$$

- If linear function f is expressed in matrix form $f(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}}$ then its kernel is the null space of the columns of the associated matrix $\underline{\underline{\boldsymbol{A}}}$

$$\mathrm{Ker}(f) = \mathrm{Null}(\underline{\underline{\boldsymbol{A}}})$$

Domain, Image, Co-domain of a Function

- Consider a (linear or not) function $f : \mathcal{V} \rightarrow \mathcal{W}$
- The **domain** of f $\mathcal{V}$ is the set of all values where the function is defined
- The **image of domain** $f(\mathcal{V})$ is the set of all values that the function can assume
- The **co-domain** $\mathcal{W}$ is the set where the function assumes its value
  - E.g., $\mathbb{R}^2$

One-to-one Function

- Consider a function $f : \mathcal{V} \to \mathcal{W}$
- A function is **one-to-one** (or injective) iff any two different elements $v_1 \neq v_2 \in \mathcal{V}$ have different images $f(v_1) \neq f(v_2)$
- Equivalently:
  - Using contrapositive: if two elements v_1 and v_2 have the same image $f(v_1) = f(v_2)$, then they are equal $v_1 = v_2$
  - In terms of set cardinality: $|f(\mathcal{V})| = |\mathcal{V}|$, i.e., the image of the domain has the same number of elements as the domain
- A *linear* function is one-to-one iff its kernel is the trivial vector space, $\text{Ker}(f) = \{\underline{\mathbf{0}}\}$
  - Equivalently the associated matrix has $\text{Null}(\underline{\underline{\boldsymbol{A}}}) = \{\underline{\mathbf{0}}\}$

Onto function

- Consider a function $f : \mathcal{V} \to \mathcal{W}$
- A function is **onto** (or surjective) iff for any element of its co-domain $w \in \mathcal{W}$, there exists an element of the domain $v \in \mathcal{V}$ that is transformed into it, i.e., $f(v) = w$
- Equivalently in terms of set cardinality:
  - $f(\mathcal{V}) = \mathcal{W}$, i.e., the image of the domain is equal to the co-domain
- Any function can be made surjective by restricting $\mathcal{W}$ to $f(\mathcal{V})$

Invertible function

- A function $f : \mathcal{V} \to \mathcal{W}$ is invertible iff it is both one-to-one (injective) and onto (surjective), i.e.,

$$\forall w \in W \quad \exists! v \in V : f(v) = w$$

- Equivalently in terms of set cardinality:
  - $|\mathcal{V}| = |\mathcal{W}|$, i.e., the co-domain and the domain have the same number of elements
- Consider an invertible function $f : \mathcal{V} \to \mathcal{W}$, the inverse of f is:
  - $f^{-1} : \mathcal{W} \to \mathcal{V}$
  - $f \circ f^{-1}$ is the identity function

Linear function composition in matrix terms

- There is a correspondence between linear functions and matrices
- Consider two matrices $\underline{\underline{\boldsymbol{A}}}$ and $\underline{\underline{\boldsymbol{B}}}$ and the two associated functions:
  - $f(\underline{\boldsymbol{y}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{y}}$
  - $g(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{B}}} \cdot \underline{\boldsymbol{x}}$
- The composed function is defined as:

$$h(\underline{\boldsymbol{x}}) = (f \circ g)(\underline{\boldsymbol{x}}) = f(g(\underline{\boldsymbol{x}}))$$

- It can be shown that the associated matrix to the composed function is

$$h(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\underline{\boldsymbol{B}}} \cdot \underline{\boldsymbol{x}}$$

Matrix inverse

- Using the definition of inverse functions, two square matrices $\underline{\underline{\boldsymbol{A}}}$ and $\underline{\underline{\boldsymbol{B}}}$ are inverses iff $\underline{\underline{\boldsymbol{A}}} \cdot \underline{\underline{\boldsymbol{B}}} = \underline{\underline{\boldsymbol{I}}}$
  - We indicate the (unique) inverse of $\underline{\underline{\boldsymbol{A}}}$ with $\underline{\underline{\boldsymbol{A}}}^{-1}$
- Given an invertible (square) matrix $\underline{\underline{\boldsymbol{A}}}$, then the associated function $f(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}}$ is an invertible function, i.e., $f()$ is one-to-one and onto
- The matrix-vector equation $\underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}} = \underline{\boldsymbol{b}}$ has one and only one solution $\underline{\boldsymbol{x}}$ for any $\underline{\boldsymbol{b}}$, i.e., $\underline{\underline{\boldsymbol{A}}}^{-1}\underline{\boldsymbol{b}}$
- If $\underline{\underline{\boldsymbol{A}}}$ and $\underline{\underline{\boldsymbol{B}}}$ are invertible and can be multiplied, then $\underline{\underline{\boldsymbol{A}}} \cdot \underline{\underline{\boldsymbol{B}}}$ is invertible

Linear dependence

- Linear algebra
    - Vector and vector spaces
    - Affine spaces
    - Linear functions
    - **Linear dependence**
    - Basis
    - Dimension of a vector space
    - Direct sum
    - Connections between Machine Learning and Linear Algebra

Coordinate representation of a vector

- Consider:
  - $\mathcal{V}$ a vector space
    - $\mathcal{V}$ is not necessarily like $\mathbb{F}^n$, i.e., the vector space does not need to be a "numeric" vector representation
  - Vectors $\underline{\boldsymbol{a}}_1, \ldots, \underline{\boldsymbol{a}}_n \in \mathcal{V}$
  - $\underline{\boldsymbol{v}} \in \mathrm{Span}(\underline{\boldsymbol{a}}_1, \ldots, \underline{\boldsymbol{a}}_n) \in \mathcal{V}$
- You can represent $\underline{\boldsymbol{v}}$ in terms of $\underline{\boldsymbol{a}}_1, \ldots, \underline{\boldsymbol{a}}_n$ using the coordinates $\underline{\boldsymbol{u}} = (\alpha_1, \ldots, \alpha_n)$ such that:

$$\underline{\boldsymbol{v}} = \sum_{i=1}^{n} \alpha_i \underline{\boldsymbol{a}}_i$$

  or in terms of the matrix $\underline{\underline{\boldsymbol{A}}}$ that has $\underline{\boldsymbol{a}}_i$ as columns: $\underline{\boldsymbol{v}} = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{u}}$
- A generator set allows to represent any vector in terms of coordinates (although not unique)
- To find the coordinates $\underline{\boldsymbol{u}}$ of a vector $\underline{\boldsymbol{v}}$ with respect to a generator $\underline{\underline{\boldsymbol{A}}}$, solve the matrix-vector equation for $\underline{\boldsymbol{u}}$:

$$\underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{u}} = \underline{\boldsymbol{v}}$$

Linear dependence between vectors: definition

- A set S of n vectors $\underline{\boldsymbol{v}}_1, ..., \underline{\boldsymbol{v}}_n$ are **linearly dependent** *iff*

$$\exists(\alpha_1, ..., \alpha_n) \neq \underline{\mathbf{0}}_n : \alpha_1 \underline{\boldsymbol{v}}_1 + ... + \alpha_n \underline{\boldsymbol{v}}_n = \underline{\mathbf{0}}_k$$

- In words, the zero vector can be written as a *non-trivial* linear combination of the vectors
- Equivalently: at least one vector of the set S can be expressed as linear combination of the remaining using coefficients not all zero
- **Remarks**
  - One of the vectors $\underline{\boldsymbol{v}}_i$ can be the zero vector
  - The notion of linear dependence / independence applies to *set of vectors* and not to a single vector
    - The zero vector $\underline{\mathbf{0}}$ is not "linearly dependent", rather the set containing only the zero vector $\{\underline{\mathbf{0}}\}$ is linearly dependent
  - A *trivial* linear combination of the vectors $\underline{\boldsymbol{x}}_1, \underline{\boldsymbol{x}}_2, ..., \underline{\boldsymbol{x}}_n$ is a combination of vectors with coefficients all equal to zero
    - It is always equal to the zero vector $\underline{\mathbf{0}}$
- E.g., $S = \{(1,0,0), (0,2,0), (2,4,0)\}$ are linearly dependent since

$$2 \cdot (1,0,0) + 2 \cdot (0,2,0) - 1 \cdot (2,4,0) = (0,0,0)$$

Linear independence: definition

- Linear independence is the negation of the definition of linear dependence:

$$\neg(\exists \underline{\boldsymbol{\alpha}} \neq \underline{\mathbf{0}} : \sum \alpha_i \underline{\boldsymbol{v}}_i = \underline{\mathbf{0}})$$

  - By negating the existential quantifier: $\neg(\exists x : P(x)) \iff \forall x : \neg P(x)$

$$\forall \underline{\boldsymbol{\alpha}} \neq \underline{\mathbf{0}} \quad \sum \alpha_i \underline{\boldsymbol{v}}_i \neq \underline{\mathbf{0}}$$

- Equivalently by contrapositive $(P \implies Q) \iff (\neg Q \implies \neg P)$

$$\sum \alpha_i \underline{\boldsymbol{v}}_i = \underline{\mathbf{0}} \implies \underline{\boldsymbol{\alpha}} = \underline{\mathbf{0}}$$

  - In words, the only combination of the vectors that gives the zero vector is the trivial linear combination
- E.g., $(1, 0, 0), (0, 2, 0), (0, 0, 3)$ are linearly independent since:

$$\begin{aligned}
&\alpha_1 \cdot (1, 0, 0) + \alpha_2 \cdot (0, 2, 0) + \alpha_3 \cdot (0, 0, 3) \\
&= (\alpha_1, 2\alpha_2, 3\alpha_3) \\
&= (0, 0, 0) \\
&\iff \alpha_1 = \alpha_2 = \alpha_3 = 0
\end{aligned}$$

Property of linear dependence / independence

- Any set of vectors that contains the $\underline{\mathbf{0}}$ is linearly dependent
  - In fact we can make a non-trivial combination of the zero setting the coefficient of the zero vector non zero and all remaining zero
  - Thus the set $\{\underline{\mathbf{0}}\}$ is not linearly independent
- A subset of linearly independent vectors is still independent
- A superset of linearly dependent vectors is still dependent
  - This is the contrapositive of the previous proposition

Linear dependence and null space

- By definition linear dependence of vectors $\underline{\boldsymbol{v}}_1, \ldots, \underline{\boldsymbol{v}}_n$ means that there is a non-trivial combination of those vectors that is equal to zero vector
- If $\underline{\underline{\boldsymbol{V}}}$ is the matrix with $\underline{\boldsymbol{v}}_i$ as columns, linear dependence is written:

$$\exists \underline{\boldsymbol{x}} \neq \underline{\mathbf{0}} : \underline{\underline{\boldsymbol{V}}} \cdot \underline{\boldsymbol{x}} = \underline{\mathbf{0}}$$

- In matrix form

$$\mathrm{Null}(\underline{\underline{\boldsymbol{V}}}) \neq \{\underline{\mathbf{0}}\}$$

  i.e., the null space of columns of $\underline{\underline{\boldsymbol{V}}}$ is not trivial
- Linear independence is equivalent to $\mathrm{Null}(\underline{\underline{\boldsymbol{V}}}) = \{\underline{\mathbf{0}}\}$ i.e., the null space of the columns of $\underline{\underline{\boldsymbol{V}}}$ is the trivial space

Linear one-to-one function and linear independence

- A linear one-to-one function f preserve linear independence
- It can be proven by showing that a linear combination of transformed independent vectors is $\underline{\mathbf{0}}$ only if it is a trivial combination:

$$\underline{\boldsymbol{x}}_i \text{ independents}$$
$$\sum \alpha_i f(\underline{\boldsymbol{x}}_i) = \underline{\mathbf{0}} \implies \underline{\boldsymbol{\alpha}} = \underline{\mathbf{0}}$$

Superfluous-vector lemma

- Given a set of vectors S
- We already know that vectors are linearly dependent *iff* a vector $\underline{\boldsymbol{v}} \in S$ is a non-trivial linear combination of the remaining vectors
- This is equivalent to:

$$\exists \underline{\boldsymbol{v}} \in S : \mathrm{Span}(S) = \mathrm{Span}(S - \{\underline{\boldsymbol{v}}\})$$

  - In words, one can always remove at least a vector dependent from a generator set without changing its span
- **Intuition**: it establishes a relationship between linear dependence and span
- **Corollary**:
  - You can always *add* or *remove* linearly dependent vectors from a set, without changing its span

Equivalent questions in linear algebra

- How can we tell if:
  - Vectors $\underline{\boldsymbol{v}}_1, \ldots, \underline{\boldsymbol{v}}_n$ are linearly dependent?
  - The null space of a matrix is trivial?
  - The solution set of a homogeneous linear system is trivial?
  - A given solution of a non-homogeneous linear system is the only solution?

Basis

- Linear algebra
  - Vector and vector spaces
  - Affine spaces
  - Linear functions
  - Linear dependence
  - **Basis**
  - Dimension of a vector space
  - Direct sum
  - Connections between Machine Learning and Linear Algebra

Basis of a vector space: definition

- A set of vectors B is a **basis** of the vector space $\mathcal{V}$ iff by definition:
  - B is a set of generators for $\mathcal{V}$, i.e., $\text{Span}(B) = \mathcal{V}$
  - Vectors in B are linearly independent
- In words, a basis is a set of linearly independent generators
- Often you care about the order of the vectors
  - You should use B as a tuple or a matrix, and not a set of vectors

Unique-representation lemma

- Consider:
  - $\mathcal{V}$ is a vector space over a scalar field $\mathbb{F}$
  - $B = (\underline{\boldsymbol{b}}_1, \ldots, \underline{\boldsymbol{b}}_n)$ be a basis for $\mathcal{V}$
- For any vector $\underline{\boldsymbol{v}} \in \mathcal{V}$ there is exactly one representation of $\underline{\boldsymbol{v}}$ in terms of the basis vectors
- **Intuition**: a vector space can always be represented in terms of coordinate vectors $\mathbb{F}^n$ using a basis
- There is a function $f : \mathbb{F}^n \to \mathcal{V}$ that:
  - transforms the coordinates $\underline{\boldsymbol{x}}$ with respect to a basis $\underline{\underline{\boldsymbol{B}}}$ into a vector $\underline{\boldsymbol{v}} = \underline{\underline{\boldsymbol{B}}} \cdot \underline{\boldsymbol{x}} \in \mathcal{V}$
  - is one-to-one (because of linear independence of basis vectors and linearity of function); and
  - is onto (because a generator)
- There is also a function $g : \mathcal{V} \to \mathbb{F}^n$ that
  - Computes the (unique) coordinates of $\underline{\boldsymbol{v}} \in \mathcal{V}$ with respect to the basis B
  - f and g are inverse of each other

Change of basis transformation

- Consider two bases for the same vector space $\mathcal{V}$
  - $\underline{\boldsymbol{a}}_1, \dots, \underline{\boldsymbol{a}}_n$
  - $\underline{\boldsymbol{c}}_1, \dots, \underline{\boldsymbol{c}}_k$
  - We assume $n \neq k$ (we don't know yet that all bases of a vector space must have the same cardinality)
- We know that:
  - The matrix $\underline{\underline{\boldsymbol{A}}}$ with $\underline{\boldsymbol{a}}_i$ as columns transforms the coordinates with respect to to $\underline{\boldsymbol{a}}_i$ into a vector in $\mathcal{V}$
  - The matrix $\underline{\underline{\boldsymbol{C}}}^{-1}$ transforms a vector in $\mathcal{V}$ into its coordinates with respect to to $\underline{\boldsymbol{c}}_i$
- We can compose the 2 transformations to get a transformation from the coordinate space $\underline{\boldsymbol{a}}$ into the coordinate space $\underline{\boldsymbol{c}}$

$$\underline{\boldsymbol{c}} = \underline{\underline{\boldsymbol{C}}}^{-1} \cdot \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{a}}$$

  - This is called "a change of basis transformation"
  - The entire transformation is also invertible, i.e., we can go from coordinates of $\underline{\boldsymbol{c}}$ to $\underline{\boldsymbol{a}}$

Dimension of a vector space

- Linear algebra
  - Vector and vector spaces
  - Affine spaces
  - Linear functions
  - Linear dependence
  - Basis
  - **Dimension of a vector space**
  - Direct sum
  - Connections between Machine Learning and Linear Algebra

Simplified exchange lemma

- Consider:
  - A set S of vectors, not necessarily linearly independent
  - $\underline{\boldsymbol{z}} \neq \underline{\mathbf{0}} \in \text{Span}(S)$
- There is a vector $\underline{\boldsymbol{w}} \in S$ such that:

$$\text{Span}(S \cup \{\underline{\boldsymbol{z}}\} - \{\underline{\boldsymbol{w}}\}) = \text{Span}(S)$$

  - In words, one can exchange a vector from S with a vector from its span without changing the span
- Example
  - $S = \{\underline{\boldsymbol{v}}_1 = (1,0,0), \underline{\boldsymbol{v}}_2 = (0,1,0), \underline{\boldsymbol{v}}_3 = (1,2,0)\}$ and $\underline{\boldsymbol{z}} = (-1,-1,0)$
  - Note that $\underline{\boldsymbol{z}} = -1 \cdot \underline{\boldsymbol{v}}_1 - 1 \cdot \underline{\boldsymbol{v}}_2$
  - WLOG we can express $\underline{\boldsymbol{v}}_1$ in terms of $\underline{\boldsymbol{z}}, \underline{\boldsymbol{v}}_2$
  - Thus we can exchange $\underline{\boldsymbol{z}}$ with $\underline{\boldsymbol{v}}_1$ without changing the span of S

Exchange lemma

- Consider
  - A set S of vectors
  - A subset of *linearly independent* $A \subseteq S$ (aka "protected subset" of S)
  - Pick $\underline{\boldsymbol{z}} \in \text{Span}(S)$ such that $A \cup \{\underline{\boldsymbol{z}}\}$ is linearly independent (i.e., $\underline{\boldsymbol{z}} \notin \text{Span}(A)$)
- There is a vector $\underline{\boldsymbol{w}} \in S - A$ such that:

$$\text{Span}(S \cup \{\underline{\boldsymbol{z}}\} - \{\underline{\boldsymbol{w}}\}) = \text{Span}(S)$$

- In words, we can exchange $\underline{\boldsymbol{z}}$ with a vector of the set, without changing the span and the protected subset

Morphing lemma

- Consider
  - A vector space $\mathcal{V}$
  - G be a set of generators for $\mathcal{V}$
  - B be a basis for $\mathcal{V}$
- $|B| \leq |G|$, i.e., the cardinality of a basis is always smaller than the cardinality of a generator
- It is called morphing lemma since you can morph G into B without changing the span

A basis as smallest generator set

- Using the morphing lemma, a basis for $\mathcal{V}$ is the smallest generating set for $\mathcal{V}$
- Subset-basis lemma: any *finite* generator set S includes a basis B for $\mathcal{V} = \mathrm{Span}(S)$
- All bases for $\mathcal{V}$ have the same size

Dimension of vector space

- The dimension of a vector space $\mathcal{V}$, written $\text{Dim}(\mathcal{V})$, is the size of a (any) basis for $\mathcal{V}$
- By definition of basis, $\text{Dim}(\mathcal{V})$ is the (exact) number of linearly independent vectors generating a vector space

Rank of a set S of vectors / matrix

- Rank is the dimension of the space spanned by a set:

$$\mathrm{Rank}(S) \triangleq \mathrm{Dim}(\mathrm{Span}(S))$$

- By definition of basis, the rank of a set of vectors S is the number of linearly independent vectors included in S
- MEM: Rank and Dim measure the same thing, but for vectors and vector spaces

Rank and dimension: example

- The vectors $S = \{(1,0,0),(0,2,0),(2,4,0)\}$ are linearly dependent, thus Rank(S) is < 3
- Actually Rank(S) $= 2$ since the first 2 vectors are linearly independent (basis)

Nullity of a set of vectors / matrix

- Given a set S of vectors, or column vectors in a matrix
- Nullity is the dimension of the null space of S, i.e., $\mathrm{Dim}(\mathrm{Null}(S))$
- It is the dual of Rank

Dimension of trivial vector space

- Consider the vector space $\mathcal{V} = \text{Span}(\{\underline{\mathbf{0}}\})$
- One could think that the dimension is 1 since it is spanned by 1 vector
- This is not possible since the set with the zero vector is not independent and thus cannot be a basis
- In reality $\mathcal{V}$ is spanned by the empty set, thus $\text{Dim}(\mathcal{V}) = 0$

Subspace dimension lemma

- $\mathcal{U}$ is a vector subspace of $\mathcal{W}$
- ***Thesis***

  $$\mathrm{Dim}(\mathcal{U}) \leq \mathrm{Dim}(\mathcal{W})$$

  and

  $$\mathrm{Dim}(\mathcal{U}) = \mathrm{Dim}(\mathcal{W}) \iff \mathcal{U} = \mathcal{W}$$

- ***Proof***
- One can use a modified version of the exchange lemma to have basis of $\mathcal{W}$ include a basis of $\mathcal{U}$

Subspace dimension lemma: example

- $\mathcal{V} = \text{Span}\{(1,2),(2,1)\}$
- Since $\mathcal{V} \subseteq \mathbb{R}^2$, then $\text{Dim}(\mathcal{V}) \leq 2$
- Since the vectors are independent then $\text{Dim}(V) = 2$ and $\mathcal{V} = \mathbb{R}^2$

Column space of a matrix

- $\text{Col}(\underline{\underline{\boldsymbol{A}}}) = \text{Span}(\text{columns of } \underline{\underline{\boldsymbol{A}}})$

Row space of a matrix

- $\mathrm{Row}(\underline{\underline{\boldsymbol{A}}}) = \mathrm{Span}(\text{rows of } \underline{\underline{\boldsymbol{A}}}) = \mathrm{Span}(\underline{\underline{\boldsymbol{A}}}^T) = \mathrm{Col}(\underline{\underline{\boldsymbol{A}}}^T)$

Row / column rank of a matrix $\underline{\underline{\boldsymbol{M}}}$

- $=$ rank of (i.e., dimension of the space spanned by) row / column vectors of $\underline{\underline{\boldsymbol{M}}}$
- We will see that these ranks are the same, so we can talk of "rank of a matrix"

Row / column rank of a matrix: example

- The matrix
$$\underline{\underline{\boldsymbol{M}}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 2 & 4 & 0 \end{pmatrix}$$
has both row and column rank equal to 2
- The matrix
$$\underline{\underline{\boldsymbol{M}}} = \begin{pmatrix} 1 & 0 & 0 & 5 \\ 0 & 2 & 0 & 7 \\ 0 & 0 & 3 & 9 \end{pmatrix}$$
has both row and column rank equal to 3

Matrix rank theorem: lemma

- For any matrix A

$$\mathrm{Rank}(\mathrm{Row}(\underline{\underline{\boldsymbol{A}}})) \leq \mathrm{Rank}(\mathrm{Col}(\underline{\underline{\boldsymbol{A}}}))$$

Matrix rank theorem: lemma proof

- Assume that $\underline{\underline{\boldsymbol{A}}}$ is $m \times n$
- Consider the column space $\underline{\underline{\boldsymbol{A}}} = (\underline{\boldsymbol{a}}_1|\underline{\boldsymbol{a}}_2|...|\underline{\boldsymbol{a}}_n)$ which is generated by n vectors with m components
- This space has a dimension $k = \mathrm{Rank}(\mathrm{Col}(\underline{\underline{\boldsymbol{A}}})) \leq n$ and let's call the basis $\underline{\underline{\boldsymbol{B}}}$
- So each $\underline{\boldsymbol{a}}_i = \alpha_1 \underline{\boldsymbol{b}}_1 + ... + \alpha_k \underline{\boldsymbol{b}}_k = \underline{\underline{\boldsymbol{B}}} \cdot \underline{\boldsymbol{\alpha}}$ where the $\underline{\boldsymbol{b}}_j$ has m components
- If we write all vectors $\underline{\boldsymbol{a}}_i = \underline{\underline{\boldsymbol{A}}}\underline{\boldsymbol{\alpha}}_i$ and use the matrix-matrix product in terms of matrix-vector product we get $\underline{\underline{\boldsymbol{A}}} = \underline{\underline{\boldsymbol{BC}}}$ $(m \times n = (m \times k) \cdot (k \times n))$
- If we transpose we have:

$$\underline{\underline{\boldsymbol{A}}}^T = \underline{\underline{\boldsymbol{C}}}^T \cdot \underline{\underline{\boldsymbol{B}}}^T = (n \times k) \cdot (k \times m)$$

  so we can express the rows of $\underline{\underline{\boldsymbol{A}}}^T$ (i.e., the columns of $\underline{\underline{\boldsymbol{A}}}$) as a linear combination of the k rows of $\underline{\underline{\boldsymbol{B}}}^T$ (vector-matrix product)
- So the k rows of $\underline{\underline{\boldsymbol{B}}}^T$ are a generator for rows of $\underline{\underline{\boldsymbol{A}}}$
- They contain a basis for Row($\underline{\underline{\boldsymbol{A}}}$) because of basis generator lemma:

Matrix rank theorem

- For every matrix $\underline{\underline{\boldsymbol{A}}}$, row rank equals column rank:

$$\mathrm{Rank}(\mathrm{Row}(\underline{\underline{\boldsymbol{A}}})) = \mathrm{Rank}(\mathrm{Col}(\underline{\underline{\boldsymbol{A}}}))$$

Matrix rank theorem: proof

- Penultimate step: to prove $a = b$ we can prove that $a \leq b$ and the converse $b \leq a$
- We can always exchange rows and columns of a matrix by transposing
- So we get from the lemma:

$$\mathrm{Rank}(\mathrm{Row}(\underline{\underline{\boldsymbol{A}}}^T)) \leq \mathrm{Rank}(\mathrm{Col}(\underline{\underline{\boldsymbol{A}}}^T))$$

but $\mathrm{Row}(\underline{\underline{\boldsymbol{A}}}^T) = \mathrm{Col}(\underline{\underline{\boldsymbol{A}}})$ so we have

$$\mathrm{Rank}(\mathrm{Col}(\underline{\underline{\boldsymbol{A}}})) \leq \mathrm{Rank}(\mathrm{Row}(\underline{\underline{\boldsymbol{A}}}))$$

and we can use the penultimate step to reach the thesis

Direct sum

- Linear algebra
  - Vector and vector spaces
  - Affine spaces
  - Linear functions
  - Linear dependence
  - Basis
  - Dimension of a vector space
  - **Direct sum**
  - Connections between Machine Learning and Linear Algebra

Minimal intersection of subspaces

- Given two vector spaces $\mathcal{U}$ and $\mathcal{V}$, subsets of the same vector space $\mathcal{W}$

$$\mathcal{U}, \mathcal{V} \subseteq \mathcal{W}$$

- The minimal possible intersection of two vector spaces is

$$\mathcal{U} \cap \mathcal{V} = \{\underline{\mathbf{0}}\}$$

- In other words it is not possible for two subspaces to have no intersection $calU \cap \mathcal{V} = \emptyset$

Direct sum of vector spaces: definition

- Given two vector spaces $\mathcal{U}$ and $\mathcal{V}$, subsets of the same vector space $\mathcal{W}$

$$\mathcal{U}, \mathcal{V} \subseteq \mathcal{W}$$

- Assume $\mathcal{U} \cap \mathcal{V} = \{\underline{\mathbf{0}}\}$
- We define direct sum of $\mathcal{U}$ and $\mathcal{V}$:

$$\mathcal{U} \oplus \mathcal{V} = \{\underline{\boldsymbol{u}} + \underline{\boldsymbol{v}} : \underline{\boldsymbol{u}} \in \mathcal{U}, \underline{\boldsymbol{v}} \in \mathcal{V}\}$$

- MEM: It's like a linear combination of vectors from the two sub-spaces, the coefficients α, β are provided by the underlying linearity of vector space

Direct sum: geometric interpretation

- $\mathcal{U}, \mathcal{V} \subseteq \mathbb{R}^n$ are spanned by two independent vectors (i.e., two non-parallel lines)
- The direct sum $\mathcal{U} \oplus \mathcal{V}$ is the plane containing both lines

Direct sum: properties

- $\mathcal{U} \oplus \mathcal{V}$ is a vector space (by def closed with respect to $+$ and *)
- The union of two generators (resp. bases) for $\mathcal{U}$ and for $\mathcal{V}$ is a generator (resp. basis) for $\mathcal{U} \oplus \mathcal{V}$
- Each vector has a unique representation as sum of a vector from $\mathcal{U}$ and $\mathcal{V}$ (by unique representation lemma)
- The dimension of $\mathcal{U} \oplus \mathcal{V}$ is the sum of the dimensions

Direct sum: example

- $\mathcal{U} = \text{Span}(1000, 0100)$ and $\mathcal{V} = \text{Span}(0010)$ over GF(2)
- The vectors have no intersection besides $\underline{\mathbf{0}}$ since their bases have no intersection
- $\mathcal{U} \oplus \mathcal{V} = \{0000, 1000, 0100, 0010, 1100, 1010, 0110, 1110\} = \text{Span}(1000, 0100, 0010)$

Complementary subspace

- Two spaces $\mathcal{U}, \mathcal{V} \subseteq \mathcal{W}$ are complementary subspaces of $\mathcal{W} \iff$ their direct sum is $\mathcal{U} \oplus \mathcal{V} = \mathcal{W}$
- Thus, any vector in $\mathcal{W}$ can be written as the (unique) sum of a vector of $\mathcal{U}$ and a vector of $\mathcal{V}$

Existence of complementary subspaces

- Given a space $\mathcal{W}$ and its subspace $\mathcal{U} \subseteq \mathcal{W}$, exists one (unique) subspace $\mathcal{V}$ such that $\mathcal{U}, \mathcal{V}$ are complementary subspaces
- It can be proved using basis of $\mathcal{U}$ and $\mathcal{V}$

Connections between Machine Learning and Linear Algebra

- Linear algebra
  - Vector and vector spaces
  - Affine spaces
  - Linear functions
  - Linear dependence
  - Basis
  - Dimension of a vector space
  - Direct sum
  - **Connections between Machine Learning and Linear Algebra**