LINEAR ALGEBRA

Introductory Notes for Machine Learning

Michael Hoon

December 6, 2023

Contents

1 Introduction **3**

1.1 Rings . . . 3

1.2 Fields . . . 4

1.3 Real Euclidean Space . . . 4

2 Vectors **4**

2.1 Vector Operations . . . 4

2.1.1 Addition . . . 4

2.1.2 Scalar Multiplication . . . 5

2.1.3 Linear Combination . . . 5

2.2 Dot Product . . . 5

2.3 Normed Spaces . . . 5

2.3.1 L^p Norms . . . 6

2.3.2 L^∞ Norm . . . 6

2.4 Unit Vectors . . . 6

2.5 Orthogonality . . . 7

2.5.1 Orthogonal Projection . . . 7

2.6 Cauchy-Schwarz Inequality . . . 7

3 Matrices **8**

3.1 Addition and Scalar Multiplication . . . 8

3.2 Transpose . . . 9

3.3 Matrix Multiplication . . . 9

3.3.1 Matrix Multiplication Properties . . . 9

3.3.2 Dot Product Revisited . . . 10

3.4 Identity Matrices . . . 10

4 Linear Mappings **11**

4.1 Isomorphisms & Homomorphisms . . . 12

5 Vector Spaces **12**

5.1 Metric Spaces . . . 12

5.2 Normed Spaces . . . 13

5.3 Inner Product Space . . . 13

6 Eigenthings **13**

7 Ordinary Least Squares **13**

1 Introduction

Linear Algebra is foremost the study of vector spaces, and the functions between vector spaces called mappings. However, underlying every vector space is a structure known as a field, and underlying every field there is what is known as a ring. We begin with the definition of a ring and proceed from there.

1.1 Rings

Definition 1.1. A ***ring*** is a triple $(R, +, \cdot)$ consisting of a set R of objects, along with binary operations ***addition*** $+ : R \times R \to R$ and ***multiplication*** $\cdot : R \times R \to R$ subject to the following axioms:

1. $a + b = b + a \; \forall a, b \in R$.
2. $a + (b + c) = (a + b) + c, \; \forall a, b, c \in R$.
3. $\exists \, 0 \in R$ s.t. $a + 0 = a, \; \forall a \in R$.
4. For each $a \in R, \; \exists - a \in R$ s.t. $- a + a = 0$. (Additive Identity)
5. $a \cdot (b \cdot c) = (a \cdot b) \cdot c, \; \forall a, b, c \in R$.
6. $a \cdot (b + c) = (a \cdot b) + (a \cdot c), \; \forall a, b, c \in R$.

We define a **subtraction** operation as follows:

$$a - b = a + (-b)$$

where

$$a - a = 0$$

Theorem 1.2. A ring $(R, +, \cdot)$ is ***commutative*** if it satisfies the additional axiom: $a \cdot b = b \cdot a, \; \forall \, a, b \in R$

Theorem 1.3. A *commutative ring* $(R, +, \cdot)$ is a ***unitary commutative ring*** if it satisfies the additional axiom: $\exists \, 1 \in R$ s.t. $a \cdot 1 = a, \; \forall \, a \in R$

Theorem 1.4. et $(R, +, \cdot)$ be a unitary ring (satisfies 1.3 but not 1.2). The ***multiplicative inverse*** of an object $a \in R$ is an object $a^{-1} \in R$ for which: $a \cdot a^{-1} = a^{-1} \cdot a = 1$

Now we have the necessary definitions to properly define a field:

1.2 Fields

Definition 1.5. A ***field*** is a unitary commutative ring $(R, +, \cdot)$ for which $1 \neq 0$, and every $a \in R \text{s.t.} a \neq 0$ has a multiplicative inverse. To summarise, a field is a set of objects $\mathbb{F}$, together with binary operations $+$ and $\cdot$ on $\mathbb{F}$, that are subject to the following ***field axioms***:

1. $a + b = b + a \ \forall\ a, b \in \mathbb{F}$.
2. $a + (b + c) = (a + b) + c,\ \forall a, b, c \in \mathbb{F}$.
3. $\exists\ 0 \in \mathbb{F}$ s.t. $a + 0 = a\ \forall\ a, b, c \in \mathbb{F}$.
4. For each $a \in \mathbb{F},\ \exists\ -a \in \mathbb{F}$ s.t. $-a + a = 0$.
5. $a \cdot (b \cdot c) = (a \cdot b) \cdot c\ \forall\ a, b, c \in \mathbb{F}$.
6. $a \cdot (b + c) = (a \cdot b) + (a \cdot c),\ \forall\ a, b, c \in \mathbb{F}$.
7. $a \cdot b = b \cdot a,\ \forall\ a, b \in \mathbb{F}$.
8. $\exists\ 0 \neq 1 \in \mathbb{F}$ s.t. $a \cdot 1 = a\ \forall\ a \in \mathbb{F}$.
9. For each $0 \neq a \in \mathbb{F}\ \exists\ a^{-1} \in \mathbb{F}$ s.t. $aa^{-1} = 1$.

where $\mathbb{F}$ denotes both the sets $\mathbb{R}$ and $\mathbb{C}$, which are commonly encountered fields in Linear Algebra.

1.3 Real Euclidean Space

Let $\mathbb{R}$ denote the set of real numbers. GIven a positive integer n, we define ***n-space*** to be the set:

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R} \text{ for } 1 \leq i \leq n\}. \tag{1}$$

where any ordered list of n objects is called an ***n-tuple***.

2 Vectors

An ordered set of numbers. For example, $\vec{v} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} \in \mathbb{R}^3$ is a 3-dimensional vector.

2.1 Vector Operations

2.1.1 Addition

Suppose $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, v_2 = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \implies \vec{v} + \vec{w} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{bmatrix}$

2.1.2 Scalar Multiplication

$\forall c \in \mathbb{R} : c \cdot \vec{v} = \begin{bmatrix} c \cdot v_1 \\ c \cdot v_2 \\ \vdots \\ c \cdot v_n \end{bmatrix}$, where c is a scalar.

2.1.3 Linear Combination

Now we combine addition with scalar multiplication to produce a **"linear combination"** of $\boldsymbol{v}$ and $\boldsymbol{w}$.

Theorem 2.1. The sum of $c\boldsymbol{v}$ and $d\boldsymbol{w}$ is a linear combination of $c\boldsymbol{v} + d\boldsymbol{w}$.

2.2 Dot Product

Definition 2.2. Let $\boldsymbol{u} = [u_1, u_2, \ldots, u_n]$ and $\boldsymbol{v} = [v_1, v_2, \ldots, v_n]$ be two vectors in $\mathbb{R}^n$. Then the dot (or inner) product of $\boldsymbol{u}$ and $\boldsymbol{v}$ is the ***real*** number:

$$\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

Some properties of the dot product:

Theorem 2.3. For any vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ and scalar c:

1. $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$
2. $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$
3. $(c\boldsymbol{u}) \cdot \boldsymbol{v} = c(\boldsymbol{u} \cdot v) = \boldsymbol{u} \cdot (c\boldsymbol{v})$
4. $\boldsymbol{u} \cdot \boldsymbol{u} > 0$ if $\boldsymbol{u} \neq 0$

Theorem 2.4. Two vectors $\boldsymbol{u}, \boldsymbol{v}$ are ***orthogonal***, written $\boldsymbol{u} \perp \boldsymbol{v}$, if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

2.3 Normed Spaces

Norms generalize the notion of length from Euclidean space.

Definition 2.5. The ***Euclidean norm*** of a vector $\boldsymbol{v} \in \mathbb{R}^n$ is $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$.

If $\boldsymbol{v} = [v_1, \ldots, v_n]$, then

$$\|\boldsymbol{v}\| = \sqrt{[v_1, \ldots, v_n] \cdot [v_1, \ldots, v_n]} = \sqrt{\sum_{i=1}^{n} v_i^2} \tag{2}$$

Which is also commonly known as the vector's magnitude or length.

Theorem 2.6. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. The ***distance*** $d(\boldsymbol{x}, \boldsymbol{y})$, between $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$$

Thus again, if $\boldsymbol{x} = [x_1, \ldots, x_n]$, $\boldsymbol{y} = [y_1, \ldots, y_n]$, then

2.3.1 L^p Norms

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Definition 2.7. Formally, the L^p norm is given by

$$\|x\|^p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}, \quad \forall p \in \mathbb{R}, p \geq 1. \tag{3}$$

Norms, including the L^p norm, are functions mapping vectors to non-negative values. On an intuitive level, the norm of a vector $\boldsymbol{x}$ measures the distance from the origin to the point $\boldsymbol{x}$.

More rigorously, a norm is any function f that satisfies the following properties:

- Positive Definite-ness: $f(\boldsymbol{x}) = 0 \implies \boldsymbol{x} = \boldsymbol{0}$
- Triangle Inequality: $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$
- Absolute homogeneity: $\forall \alpha \in \mathbb{R}, f(\alpha \boldsymbol{x} = |\alpha| f(\boldsymbol{x}))$

A vector space endowed with a norm is called a normed vector space, or simply a normed space.

2.3.2 L^∞ Norm

One other norm that commonly arises in Machine Learning is the L^∞ norm, also known as the **max norm** (the limit of the p-norm when p tends to infinity). This norm simplifies to the absolute value of the element with the largest magnitude in the vector:

$$\|\boldsymbol{x}\|_\infty = \max_i |x_i|$$

2.4 Unit Vectors

Definition 2.8. A unit vector $\boldsymbol{u}$ is a vector whose length equals 1, and $\boldsymbol{u} \cdot \boldsymbol{u} = 1$. To obtain a unit vector in the direction of any nonzero vector $\boldsymbol{v}$, divide by its length $\|\boldsymbol{v}\|$.

$$\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$$

is a unit vector in the same direction as $\boldsymbol{v}$.

2.5 Orthogonality

Given Definition 2.4, suppose $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ are orthogonal vectors. Geometrically, $\|\boldsymbol{u}+\boldsymbol{v}\|$ is the length of the longest side of the triangle, and $\|\boldsymbol{u}\|, \|\boldsymbol{v}\|$ are the shorter sides. We can see that

Theorem 2.9. Magnitude of perpendicular vectors:

$$\begin{aligned}\|\boldsymbol{u}+\boldsymbol{v}\| &= \left(\sqrt{(\boldsymbol{u}+\boldsymbol{v})\cdot(\boldsymbol{u}+\boldsymbol{v})}\right)^2 = (\boldsymbol{u}+\boldsymbol{v})\cdot(\boldsymbol{u}+\boldsymbol{v}) \\ &= (\boldsymbol{u}+\boldsymbol{v})\cdot\boldsymbol{u} + (\boldsymbol{u}+\boldsymbol{v})\cdot\boldsymbol{v} = \boldsymbol{u}\cdot\boldsymbol{v} + \boldsymbol{v}\cdot\boldsymbol{u} + \boldsymbol{u}\cdot\boldsymbol{v} + \boldsymbol{v}\cdot\boldsymbol{v} \\ &= \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2\end{aligned}$$

Which obeys the Pythagorean Theorem, and so it must be that the triangle is a right triangle (i.e. $\boldsymbol{u}, \boldsymbol{v}$ are 'perpendicular').

2.5.1 Orthogonal Projection

Theorem 2.10. Let $v \neq 0$. The orthogonal projection of $\boldsymbol{u}$ onto $\boldsymbol{v}$, $\mathrm{proj}_{\boldsymbol{v}}\boldsymbol{u}$, is given by

$$\mathrm{proj}_{\boldsymbol{v}}\boldsymbol{u} = \left(\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)$$

Proposition 2.11. If $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n, \boldsymbol{v} \neq 0$, and $c = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}$, then $\boldsymbol{u} - c\boldsymbol{v}$ is orthogonal to $\boldsymbol{v}$.

Proof. Taking the dot product,

$$(\boldsymbol{u}-c\boldsymbol{v})\cdot\boldsymbol{v} = \boldsymbol{u}\cdot\boldsymbol{v} - c(\boldsymbol{v}\cdot\boldsymbol{v}) = \boldsymbol{u}\cdot\boldsymbol{v} - \left(\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)(\boldsymbol{v}\cdot\boldsymbol{v}) = \boldsymbol{u}\cdot\boldsymbol{v} - \boldsymbol{u}\cdot\boldsymbol{v} = 0$$

where $\boldsymbol{u} - c\boldsymbol{v} \perp \boldsymbol{v}$. This also applies to any scalar a. □

2.6 Cauchy-Schwarz Inequality

Definition 2.12. If $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, then $|\boldsymbol{u}\cdot\boldsymbol{v}| \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\|$.

Proof. Suppose $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. If $\boldsymbol{u} = \boldsymbol{0}$ or $\boldsymbol{v} = \boldsymbol{0}$, then:

$$|\boldsymbol{u}\cdot\boldsymbol{v}| = |0| = 0 = \|\boldsymbol{u}\|\|\boldsymbol{v}\|$$

which affirms the theorem's conclusions. So, suppose $\boldsymbol{u}, \boldsymbol{v} \neq \boldsymbol{0}$, and let $c \in \mathbb{R}$ similar to above,

$$(\boldsymbol{u}-c\boldsymbol{v})\cdot c\boldsymbol{v} = c[(\boldsymbol{u}-c\boldsymbol{v})\cdot\boldsymbol{v}] = c(0) = 0$$

by 2.5.1. Thus $\boldsymbol{u} - c\boldsymbol{v}$ and $c\boldsymbol{v}$ are orthogonal. Using 2.9, we have

$$\|\boldsymbol{u}\|^2 = \|(\boldsymbol{u}-c\boldsymbol{v}+c\boldsymbol{v})\|^2 = \|\boldsymbol{u}-c\boldsymbol{v}\|^2 + \|c\boldsymbol{v}\|^2.$$

Since $\|\boldsymbol{u} - c\boldsymbol{v}\|^2 \geq 0$, this implies that $\|c\boldsymbol{v}\|^2 \leq \|\boldsymbol{u}\|^2$. However,

$$\|c\boldsymbol{v}\|^2 = c^2\|\boldsymbol{v}\|^2 = \left(\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)^2(\boldsymbol{v}\cdot\boldsymbol{v}) = \frac{(\boldsymbol{u}\cdot\boldsymbol{v})^2}{\boldsymbol{v}\cdot\boldsymbol{v}} = \frac{(\boldsymbol{u}\cdot\boldsymbol{v})^2}{\|\boldsymbol{v}\|^2}$$

and so from $\|c\boldsymbol{v}\|^2 \leq \|\boldsymbol{u}\|^2$ we obtain

$$\frac{(\boldsymbol{u}\cdot\boldsymbol{v})^2}{\|\boldsymbol{v}\|^2} \leq \|\boldsymbol{u}\|^2$$

whence comes $(\boldsymbol{u}\cdot\boldsymbol{v})^2 \leq \|\boldsymbol{u}\|^2\|\boldsymbol{v}\|^2$. Taking square root of both sides completes the proof. □

From here, we have

$$-\|\boldsymbol{u}\|\|\boldsymbol{v}\| \leq \boldsymbol{u}\cdot\boldsymbol{v} \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\| = -1 \leq \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|} \leq 1,\ \forall \boldsymbol{u},\boldsymbol{v} \neq \mathbf{0}$$

Theorem 2.13. Let $\boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^n$ be nonzero vectors. The angle between $\boldsymbol{u},\boldsymbol{v}$ is the number $\theta \in [0,\pi]$ for which

$$\cos\theta = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|} \quad \textbf{or} \quad \boldsymbol{u}\cdot\boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\|\cos\theta$$

3 Matrices

Definition 3.1. Let $m, n \in \mathbb{N}$, and let $\mathbb{F}$ be a field. An $m \times n$ matrix over $\mathbb{F}$ is a rectangular array of elements of $\mathbb{F}$ arranged in m rows and n columns (dimensions):

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = [a_{ij}]_{m,n} \quad \forall\ a_{ij} \in \mathbb{F} \tag{4}$$

The set of all $m \times n$ matrices with entries in the field $\mathbb{F}$ will be denoted by $\mathbb{F}^{m\times n}$:

$$\mathbb{F}^{m\times n} = \{[a_{ij}]_{m,n} : a_{ij} \in \mathbb{F} \quad \forall 1 \leq i \leq m, 1 \leq j \leq n\}, \quad \text{and} \quad \mathbb{F}^n = \mathbb{F}^{n\times 1}$$

3.1 Addition and Scalar Multiplication

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.2 Transpose

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3.3 Matrix Multiplication

Definition 3.2. Let $\mathbf{A} \in \mathbb{F}^{m\times n}$ and $\mathbf{B} \in \mathbb{F}^{n\times p}$. Then the ***product*** of $\mathbf{A}$ and $\mathbf{B}$ is the matrix $\mathbf{AB} \in \mathbb{F}^{m\times p}$

$$[\mathbf{AB}]_{ij} = \sum_{k=1}^{n} [\mathbf{A}]_{ik}[\mathbf{B}]_{kj}, \quad 1 \leq i \leq m,\ 1 \leq j \leq p$$

If we let $\mathbf{A} = [a_{ij}]_{m,n}$ and $\mathbf{B} = [b_{ij}]_{n,p}$, then $\mathbf{AB} = [c_{ij}]_{m,p}$ where

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

Thus

$$\mathbf{AB} = [a_{ij}]_{m,n}[b_{ij}]_{n,p} = \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right] \tag{5}$$

Where $\mathbf{AB} \neq \mathbf{BA}$, i.e. matrix multiplication is non-commutative.

3.3.1 Matrix Multiplication Properties

Theorem 3.3. Let $\mathbf{A} \in \mathbb{F}^{m\times n}, \mathbf{B}, \mathbf{C} \in \mathbb{F}^{p\times q}, \mathbf{D} \in \mathbb{F}^{p\times q}$, and $c \in \mathbb{F}$. Then

1. $\mathbf{A}(\mathbf{cB}) = \mathbf{c}(\mathbf{AB})$
2. Distributivity: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
3. Associativity: $(\mathbf{AB})\mathbf{D} = \mathbf{A}(\mathbf{BD})$

Definition 3.4. If $\mathbf{A} \in \mathbb{F}^{n \times n}$ and $m \in \mathbb{N}$, then

$$\mathbf{A}^m = \underbrace{\mathbf{A}\mathbf{A}\ldots\mathbf{A}}_{\text{m factors}} = \prod_{k=1}^{m} \mathbf{A}$$

3.3.2 Dot Product Revisited

Matrix multiplication is *not* commutative, but the dot product between two vectors is commutative:

$$x^\top y = (x^\top y)^\top = y^\top x \tag{6}$$

by exploiting the fact that the value of such a product is a scalar and therefore equal to its own transpose.

3.4 Identity Matrices

The **Kronecker Delta** is a function $\delta_{ij} : \mathbb{Z} \times \mathbb{Z} \to \{0, 1\}$ defined as follows for integers i, j:

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

and thus the $n \times n$ identity matrix:

$$\mathbf{I}_n = [\delta_{ij}]_n = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

Definition 3.5. For any $\mathbf{A} \in \mathbb{F}^{n \times n}$ we define $\mathbf{A}^0 = \mathbf{I}_n$ where I_n acts as **the** identity with respect to matrix multiplication, just as 1 is the identity with respect to multiplication of real numbers.

Where more generally, I_n is the only matrix for which

$$\mathbf{I}_n\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}, \quad \forall \mathbf{A} \in \mathbb{F}^{n \times n} \tag{7}$$

Definition 3.6. An $n \times n$ matrix $\mathbf{A}$ is **invertible** if there exists a matrix $\mathbf{B}$ s.t.

$$\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = \mathbf{I}_n$$

in which case we will call $\mathbf{B}$ the inverse of $\mathbf{A}$ and denote it by the symbol $\mathbf{A}^{-1}$. A matrix that is not invertible is said to be ***nonsingular***.

Theorem 3.7. Let $k \in \mathbb{N}$. If $\mathbf{A}_1, \dots, \mathbf{A}_k \in \mathbb{F}^{n\times n}$ are invertible, then $\mathbf{A}_1 \dots \mathbf{A}_k$ is invertible and

$$(\mathbf{A}_1 \dots \mathbf{A}_k)^{-1} = \mathbf{A}^{-1} \dots \mathbf{A}_1^{-1}$$

Proposition 3.8. If $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$

Proof. Suppose that $\mathbf{A} \in \mathbb{F}^{n\times n}$ is invertible, so that $\mathbf{A}^{-1}$ exists.

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n \implies (\mathbf{A}\mathbf{A}^{-1})^\top = \mathbf{I}_n^\top \implies (\mathbf{A}^{-1})^\top \mathbf{A}^\top = \mathbf{I}_n$$

and

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n \implies (\mathbf{A}^{-1}\mathbf{A}^\top) = \mathbf{I}_n^\top \implies \mathbf{A}^\top(\mathbf{A}^{-1})^\top = \mathbf{I}_n$$

Now,

$$(\mathbf{A}^{-1})^\top \mathbf{A}^\top = \mathbf{A}^\top(\mathbf{A}^{-1})^\top = \mathbf{I}_n$$

shows that $(\mathbf{A}^{-1})^\top$ is the inverse of $\mathbf{A}^\top$. Therefore $\mathbf{A}^\top$ is invertible, and moreover $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$. $\square$

4 Linear Mappings

As with functions in general, to say a mapping T **maps** a set X into a set Y, written $T : X \to Y$, means that T maps each object $x \in X$ yo a unique object $y \in Y$. We denote this by writing $T(x) = y$, and call X the **domain** of T and Y and the **codomain**. A little more formally a mapping T is a set of ordered pairs $(x, y) \in X \times Y$ with the property that

$$\forall x \in X[\exists y \in Y(((x, y) \in T) \wedge (\hat{y} \neq y \to (x, \hat{y}) \notin T))].$$

We call $T(x)$ the value of T at x. Given any set $A \subseteq X$, we define the image of A under T to be the set

$$T(A) = \{T(x) : x \in A\} \subseteq Y$$

with $T(x)$ in particular being called the **image** of T (or the **range** of T). To denote a **mapping** we write $T : \mathbb{R} \to \mathbb{R}$ for which $T(x) = \sqrt[3]{x} \quad \forall x \in \mathbb{R}$. '$\to$' is used for mappings between *sets*, while '$\mapsto$' is used *between elements of sets*.

Definition 4.1. A mapping $T : X \to Y$ is **injective** if

$$T(x_1) = T(x_2) \implies x_1 = x_2, \quad \forall x_1, x_2 \in X$$

Thus if $x_1 \neq x_2$, then $T(x_1) \neq T(x_2)$. A mapping $T : X \to Y$ is **surjective** if for each $y \in Y$ $\exists x \in X$ such that $T(x) = y$

If a mapping is both **injective** and **surjective**, we call it a **bijection**. A linear map to itself is called a **linear operator**.

Definition 4.2. Let V and W be vector spaces over $\mathbb{F}$. A mapping $L : V \to W$ is called a **linear mapping** if the following hold:

1. $L(\boldsymbol{u}+\boldsymbol{v}) = L(\boldsymbol{u}) + L(\boldsymbol{v}) \quad \forall \boldsymbol{u}, \boldsymbol{v} \in V$.
2. $L(c\boldsymbol{u}) = cL(\boldsymbol{u}) \forall c \in \mathbb{F}$ and $\boldsymbol{u} \in V$.

Proposition 4.3. If $L : V \to W$ is a linear mapping, then

1. $L(\mathbf{0}) = \mathbf{0}$
2. $L(-\boldsymbol{v}) = -L(\boldsymbol{v})$ for any $\boldsymbol{v} \in V$
3. For any $c_1 c_2 \ldots c_n \in \mathbb{F}$, $\boldsymbol{v}_1 \boldsymbol{v}_2 \ldots v_n \in V$:
$$L\left(\sum_{k=1}^{n} c_k \boldsymbol{v}_k\right) = \sum_{k=1}^{n} c_k L(\boldsymbol{v}_k)$$

4.1 Isomorphisms & Homomorphisms

The definition of a linear map is suited to reflect the structure of vector spaces, since it preserves a vector spaces' two main operations: **addition and scalar multiplication**. In algebraic terms, a linear map is called a **homomorphism** of vector spaces. An *invertible homomorphism* (where the inverse is also a homomorphism) is called an isomorphism.

Definition 4.4. A **bijective** linear mapping is called an **isomorphism**. If V and W are vector spaces and there exists a linear mapping $L : V \to W$ that is an isomorphism, then V and W are said to be **isomorphic** and we write $V \simeq W$.

Isomorphic spaces are truly identical in save for the symbols used to represent their elements. In fact any vector space V of dimension n can be shown to be isomorphic to $\mathbb{R}^n$.

5 Vector Spaces

5.1 Metric Spaces

Metrics generalize the notion of distance from Euclidean space (although metric spaces need not be vector spaces).

Definition 5.1. A metric on a set S is a function $d : S \times S \to \mathbb{R}$ that satisfies:

1. $d(x, y) \geq 0$ with equality **iff** $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, z) = d(x, y) + d(y, z)$ (**Triangle Inequality**)

for all $x, y, z \in S$. Metrics allow for limits to be defined for mathematical objects other than real numbers.

5.2 Normed Spaces

Refer to Section 2.3. Note that the axioms for **metrics** are satisfied under this definition and follow directly from the axioms for norms. Therefore, any normed space is also a metric space (If a normed space is complete with respect to the distance metric induced by its norm, we say that it is a **Banach space**).

5.3 Inner Product Space

Definition 5.2. An **inner product** on a real vector space V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$, satisfying:

1. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle$, with equality if and only if $\boldsymbol{x} = \boldsymbol{0}$
2. Linearity in the first slot: $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$ and $\langle \alpha\boldsymbol{x}, \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle$.
3. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$

for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}$, and $\alpha \in \mathbb{R}$.

A vector space endowed with an inner product is called an **inner product space**. Any inner product on V induces a norm on V:

$$\|x\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$$

Again, more formally, two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be **orthogonal** 2.5 if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. If two orthogonal vectors additionally have unit length ($\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$), then they are described as **orthonormal**.

Definition 5.3. The standard inner product on $\mathbb{R}^n$ is given by:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x}^\top \boldsymbol{y}$$

6 Eigenthings

7 Ordinary Least Squares

This chapter is heavily inspired by "Introductory Econometrics: A Modern Approach" by Jeffrey Wooldridge, and UC Berkeley's ENVECON C118 course.