Linear Algebra for Intelligent Systems
Vector Spaces

Andres Mendez-Vazquez

February 17, 2020

Abstract

Now, we start our journey with the important concepts coming from the Linear Algebra Theory. In this first section, we will look at the basic concepts of vector spaces and their dimensions. This will get us the first tools for handling many of the concepts that we use in intelligent systems. Finally, we will explore the application of these concepts in Machine Learning.

Contents

1 Introduction

About 4000 years ago, Babylonians knew how to solve the following kind of systems [3]:

$$\begin{aligned} ax + by &= c \\ dx + ey &= f \end{aligned}$$

where x, y are unknown. As always the first steps in any field of knowledge tend to be slow, something like 1800 year slow. Because, it is only after the death of Plato and Aristotle, that the Chinese (Nine Chapters of the Mathematical Art 200 B.C.) were able to solve 3×3 systems by working an "elimination method" similar to the one devised by Gauss, "The Prince of Mathematics," 2000 years later for general systems. But, it is only when Leibniz tried to solve systems of linear equations that the one of the first concepts of linear algebra came to be, the determinant. Finally, Gauss defined implicitly the concept of a Matrix as linear transformations in his book "Disquisitions." Furthermore, the Matrix term was finally introduced by Cayley in two papers in 1850 and 1858 respectively, which allowed him to prove the important Cayley-Hamilton Theorem. Much more exist in the history of linear algebra [3], and although, this is barely a glimpse of this rich history of what is one of the the most important tools for intelligent systems, it stresses its importance in it.

2 Fields and Vector Spaces

Given that fields (Set of Numbers) are so important to us in our calculations, like

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n =& b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n =& b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n =& b_2. \end{aligned}$$

It is clear that we would like to collect them in a compact structure that allows for simpler manipulation. Thus, we can do this if we use a n-tuple structure like the following ones

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \text{ and } A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{1}$$

Thus, we can write:

$$A\boldsymbol{x} = \boldsymbol{b} \tag{2}$$

This structures can be lumped together by the use of specific operations like the sum and multiplication [2, 4]. Making clear the need to define what is a space of structures. For this, we define first the concept of a field.

Definition 1. A field K is a set of real or complex numbers with the following properties

1. Addition is commutative, $x + y = y + x$ for all $x, y \in K$.
2. Addition is associative $x + (y + z) = (x + y) + z$ for all $x, y, z \in K$.
3. There is a unique element 0 in K such that $x + 0 = x$, for every $x \in K$.
4. For each $x \in F$ there is a unique element $(-x)$ in K such that $x+(-x) = 0$.
5. Multiplication is commutative $xy = yx$ for all $x, y \in K$.
6. Multiplication is associative $x\,(yz) = (xy)\,z$ for all $x, y, z \in K$.
7. There is a unique non-zero element 1 in K such that $x1 = x$, for every $x \in K$.
8. For each non-zero $x \in F$ there is a unique element x^{-1} or $\frac{1}{x}$ in K such that $xx^{-1} = 1$.
9. Multiplication distributes over addition, $x\,(y + z) = xy+xz$, for all $x, y, z \in K$.

This allows to define the concept of vector spaces.

Definition 2. A vector space V over the field K is a set of objects which can be added and multiplied by elements of K where the results are again an element of V. Having, the following properties:

1. Given elements $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ of V, we have $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$.
2. There is an element of V, denoted by O, such that $O + \boldsymbol{u} = \boldsymbol{u} + O = \boldsymbol{u}$ for all elements $\boldsymbol{u}$ of V.
3. Given an element $\boldsymbol{u}$ of V, there exists an element $-\boldsymbol{u}$ in V such that $\boldsymbol{u} + (-\boldsymbol{u}) = O$.
4. For all elements $\boldsymbol{u}, \boldsymbol{v}$ of V, we have $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.
5. For all elements $\boldsymbol{u}$ of V, we have $1 \cdot \boldsymbol{u} = \boldsymbol{u}$.
6. If c is a number, then $c\,(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{u} + c\boldsymbol{v}$.
7. if a, b are two numbers, then $(ab)\,\boldsymbol{v} = a\,(b\boldsymbol{v})$.

8. If a, b are two numbers, then $(a+b)\,\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$.

Therefore, the elements of a vector space will be **called vectors**.

Example 3. Let $V = K^n$ be the set of $n-$tuples of elements of K. Let

$$A = (a_1, a_2, ..., a_n) \text{ and } B = (b_1, b_2, ..., b_n)\,.$$

It is easy to prove that this is a vector space.

2.1 Subspaces

Now, we have the idea of subsets that can be vector spaces [2, 4] which can be highly useful for representing certain structures in data spaces.

Definition 4. Let V a vector space and $W \subseteq V$, thus W is a subspace if:

1. If $\boldsymbol{v}, \boldsymbol{w} \in W$, then $\boldsymbol{v} + \boldsymbol{w} \in W$.
2. If $\boldsymbol{v} \in W$ and $c \in K$, then $c\boldsymbol{v} \in W$.
3. The element $0 \in V$ is also an element of W.

Example 5. Given the vector space V of $m \times n$ matrices over $\mathbb{R}$. A subspace is the one defined as follow:

$$W = \{A \in V | \text{The last row has only zeros}\} \tag{3}$$

It is possible to obtain a compact

Theorem 6. *A non-empty subset W of V is a subspace of V if and only if for each pair of vectors $\boldsymbol{v}, \boldsymbol{w} \in W$ and each scalar $c \in K$ the vector $c\boldsymbol{v} + \boldsymbol{w} \in W$.*

Proof. Case $\Leftarrow$
Suppose that $W \neq \emptyset$ and $c\boldsymbol{v} + \boldsymbol{w} \in W$ for all vectors $\boldsymbol{v}, \boldsymbol{w} \in W$, $c \in K$. Thus, exist $\boldsymbol{\rho} \in W$ such that $(-1)\,\boldsymbol{\rho} + \boldsymbol{\rho} = 0 \in W$. Then, $c\boldsymbol{v} = c\boldsymbol{v} + 0 \in W$. Finally, if $\boldsymbol{v}, \boldsymbol{w} \in W$, then $\boldsymbol{v} + \boldsymbol{w} = (1)\,\boldsymbol{v} + \boldsymbol{w} \in W$. Thus, W is a subspace of V.
Case $\Rightarrow$ Too simple, left to you to prove it.

$\square$

Example 7. If V is any vector space,

1. V is a subspace of V.
2. The subset consisting of the zero vector alone is a subspace of V, called the **zero subspace** of V.
3. The space of polynomial functions over the field $\mathbb{R}$ is a subspace of the space of all functions from $\mathbb{R}$ into $\mathbb{R}$.

2.2 Linear Combinations

Now, a fundamental idea is going to presented next and represents one of the fundamentals ideas in linear algebra because its applications in:

1. Convex Representations in Optimization.
2. Endmember Representation in Hyperspectral Images (Fig. 1) [1].
3. Geometric Representation of addition of forces in Physics (Fig. 2)

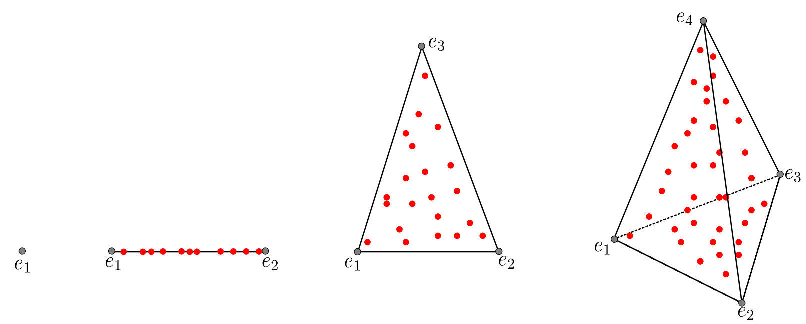

Figure 1: Mixing Model with $\boldsymbol{e} = \sum_{i=1}^{n} \alpha_i e_i$ and $\sum_{i=1}^{n} \alpha_i = 1$

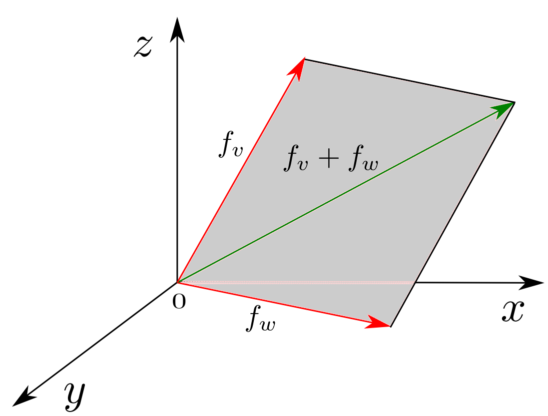

Figure 2: Forces as Vectors

Definition 8. Let V an arbitrary vector space, and let $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n \in V$ and $x_1, x_2, ..., x_n \in K$. Then, an expression like

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + ... + x_n\boldsymbol{v}_n \tag{4}$$

is called a **linear combination** of $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$.

We are ready to prove that the space W of all linear combinations of a given collection $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ is a subspace of V.

Proof. Let $y_1, y_2, ..., y_n \in K$. Then, we can get two elements in W.

$$\begin{aligned}\boldsymbol{t}_1 &= x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + ... + x_n\boldsymbol{v}_n \\ \boldsymbol{t}_2 &= y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2 + ... + y_n\boldsymbol{v}_n\end{aligned}$$

We have then that

$$\begin{aligned}\boldsymbol{t}_1 + \boldsymbol{t}_2 &= x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + ... + x_n\boldsymbol{v}_n + y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2 + ... + y_n\boldsymbol{v}_n \\ &= (x_1 + y_1)\,\boldsymbol{v}_1 + (x_2 + y_2)\,\boldsymbol{v}_2 + ... + (x_n + y_n)\,\boldsymbol{v}_n\end{aligned}$$

Thus, the sum of two elements of W is an element in W. In a similar fashion, given $c \in K$:

$$c\,(x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + ... + x_n\boldsymbol{v}_n) = cx_1\boldsymbol{v}_1 + cx_2\boldsymbol{v}_2 + ... + cx_n\boldsymbol{v}_n$$

a linear combination of the elements $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$, and an element of W. Finally,

$$0 = 0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + ... + 0\boldsymbol{v}_n$$

an element of W[1]

□

Theorem 9. *Let V be a vector space over the field K. The intersection of any collection of subspaces of V is a subspace of V.*

Proof. Let $\{W_\alpha\}_{\alpha \in I}$ be a collection of subspaces of V, and let $W = \cap_{\alpha \in I} W_\alpha$. Thus, the if $\boldsymbol{v} \in W$ then $\boldsymbol{v} \in W_\alpha\ \forall \alpha \in I$, and in addition $O \in W_\alpha\ \forall \alpha \in I$. Thus, it is easy to see that O is in all W_α, thus in the the intersection W. Now, given two vectors $\boldsymbol{v}, \boldsymbol{w} \in W$, both of them are in all W_α by definition. Then, given a $c \in K$, $c\boldsymbol{v} + \boldsymbol{w} \in W_\alpha$ for all $\alpha \in I$. Therefore, by Theorem 6, W is a subspace of V. □

This allows us to discover that there is a smallest subspace for any collection S of elements of V such that it contains S and it is contained in any subspace containing S.

Definition 10. Let S be a set of vectors in a vector space V. The **subspace spanned** by S is defined as the intersection W of all subspaces of V which contains S. When S is a finite set of vectors, $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$, we shall simply call W the **subspace spanned by the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$**.

From this, we only need to prove a simple fact about the spanned space to have an efficient tool of representation of subspaces.

[1] The subspace W is called the subspace **generated or spanned** by $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$.

Theorem 11. *The subspace spanned by $S \neq \emptyset$ of a vector space V is the set of all linear combinations of vectors in S.*

Proof. Let W be a subspace spanned by S. Then, each linear combination

$$\boldsymbol{v} = \sum_{i=1}^{n} x_i \boldsymbol{v}_i \in W$$

of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in S$ is clearly in W. Thus, W contains the set L of all linear combinations of vectors in S. Now, the set L contains S, and it is different from $\emptyset$. If $\boldsymbol{\alpha}, \boldsymbol{\beta}$ belong to the L then

$$\boldsymbol{\alpha} = \sum_{i=1}^{m} x_i \boldsymbol{v}_i,$$

$$\boldsymbol{\beta} = \sum_{i=1}^{n} y_i \boldsymbol{w}_i,$$

with $\{\boldsymbol{v}_i\}_{i=1}^{n} \subset L$ and $\{\boldsymbol{w}_i\}_{i=1}^{n} \subset L$. Thus, for any scalar c,

$$c\boldsymbol{\alpha} + \boldsymbol{\beta} = c\sum_{i=1}^{m} x_i \boldsymbol{v}_i + \sum_{i=1}^{n} y_i \boldsymbol{w}_i = \sum_{i=1}^{m} (cx_i)\, \boldsymbol{v}_i + \sum_{i=1}^{n} y_i \boldsymbol{w}_i$$

.
Therefore, $c\boldsymbol{\alpha} + \boldsymbol{\beta}$ belongs to L. Thu, L is a subspace of V by Theorem 6.

□

Example 12. Let F a subfield of the field $\mathbb{C}$ of complex numbers. Suppose

$$\boldsymbol{\alpha}_1 = (1, 2, 0, 3, 0)^T,$$
$$\boldsymbol{\alpha}_2 = (0, 0, 1, 4, 0)^T,$$
$$\boldsymbol{\alpha}_3 = (0, 0, 0, 0, 1)^T.$$

By Theorem 11, a vector $\boldsymbol{\alpha}$ is the subspace W of F^5 spanned by $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ if and only if $\exists$ scalars c_1, c_2, c_3 in F such that

$$\boldsymbol{\alpha} = c_1\boldsymbol{\alpha}_1 + c_2\boldsymbol{\alpha}_2 + c_3\boldsymbol{\alpha}_3$$

Therefore, the vectors in W have the following structure

$$\begin{aligned}\boldsymbol{\alpha} &= c_1 (1, 2, 0, 3, 0)^T + c_2 (0, 0, 1, 4, 0)^T + c_3 (0, 0, 0, 0, 1)^T \\ &= (c_1, 2c_1, c_2, 3c_1 + 4c_2, c_3)^T.\end{aligned}$$

Furthermore, W can be described as the set

$$\{(x_1, x_2, x_3, x_4 x_5)\, | x_2 = 2x_1, x_4 = 3x_1 + 4x_3\}.$$

3 Basis and Dimensions

Before we define the concept of basis, it is interesting to have a intuitive idea about why the need of a basis. It is clear that we would like to have a compact representation of specific spaces to avoid listing all the possible elements in it. For example, given the (Fig 3), we have that all the points in the straight line can be defined as

$$\{c\,(\alpha,\beta,\gamma)\,|c\in\mathbb{R}\} \tag{5}$$

Thus, it necessary to define the concepts of a linearly dependent and independent basis.

Definition 13. Let V be a vector space over a field K, and let $\boldsymbol{v}_1,\boldsymbol{v}_2,...,\boldsymbol{v}_n\in V$. We have that $\boldsymbol{v}_1,\boldsymbol{v}_2,...,\boldsymbol{v}_n$ are linearly dependent over K if there are scalars $a_1,a_2,...,a_n\in K$ not all equal to 0 such that

$$a_1\boldsymbol{v}_1+a_2\boldsymbol{v}_2+...+a_n\boldsymbol{v}_n=O$$

Therefore, if there are not such numbers, then we say that $\boldsymbol{v}_1,\boldsymbol{v}_2,...,\boldsymbol{v}_n$ are linearly independent.

Example 14. Let $V=K^n$ and consider the following vectors:

$$\begin{aligned}\boldsymbol{e}_1&=(1,0,...,0)^T\\&\vdots\\\boldsymbol{e}_n&=(0,0,...,1)^T\end{aligned}$$

The operator $\boldsymbol{v}^T$ is the exchange of the rows by columns. Then

$$a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2+...+a_n\boldsymbol{e}_n=(a_1,a_2,...,a_n) \tag{6}$$

Then, we have that $(a_1,a_2,...,a_n)=0\Leftrightarrow a_i=0$ for all $1\leq i\leq n$. Thus, $\boldsymbol{e}_1,\boldsymbol{e}_2,...,\boldsymbol{e}_n$ are linearly independent.

Example 15. Let V be the vector space of all functions $f:\mathbb{R}\rightarrow\mathbb{R}$ of a variable t. Then, we want to find a sequence of functions that are linear independent, and for this we have the following sequence of functions:

$$e^t,e^{2t},e^{3t} \tag{7}$$

Then, we have the following to prove that they are linearly independent.

$$a_1e^t+a_2e^{2t}+a_3e^{3t}=0.$$

Deriving, we get

$$a_1e^t+2a_2e^{2t}+3a_2e^{3t}=0.$$

Finally, we subtract

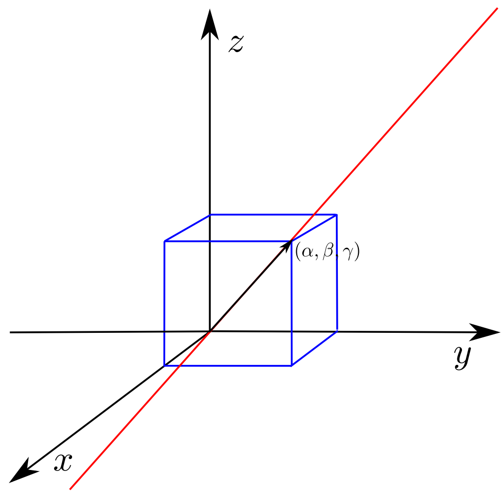

Figure 3: A basis of one element

$$a_1e^t + 2a_2e^{2t} + 3a_2e^{3t} - \left(a_1e^t + a_2e^{2t} + a_3e^{3t}\right) = a_2e^{2t} + 2a_3e^{3t} = 0.$$

We again to subtract the previous equation from $a_1e^t + a_2e^{2t} + a_3e^{3t} = 0$. Thus,

$$a_1e^t + a_2e^{2t} + a_3e^{3t} - \left(a_2e^{2t} + 2a_3e^{3t}\right) = a_1e^t - a_3e^{3t} = 0$$

In this way, we have $a_1e^t = a_3e^{3t}$ for all t. Then, $a_3e^{2t} = a_1$, and assume that $a_1 \neq 0$ and $a_3 \neq 0$.

$$e^{2t} = \frac{a_1}{a_3}.$$

But t can take any value, therefore, it is not possible. Then, the other option is $a_1 = a_3 = 0$. From this, we have that $a_2 = 0$.
Following these examples, we have the following consequences from the previous definition 13 of linear dependence and independence.

1. Any set which contains a linearly dependent set is linearly dependent.
2. Any subset of a linearly independent set is linearly independent.
3. Any set which contains the O is linearly dependent because $1 \cdot O = O$.
4. A set S of vectors is linearly independent set if and only if each finite subset of S is linearly independent.

Now, from these basic ideas comes the concept of a **basis.**

Definition 16. If elements $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ generate or span V and in addition are linearly independent, then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ is called a **basis** of V. In other words the elements $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ form a basis of V.

Thus, we can say that the vectors in (Example 14) and the ones in (Example 15) are basis for K^n and the space generated by $\left\{e^t, e^{2t}, e^{3t}\right\}$.

3.1 Coordinates

Now, we need to define the concept of coordinates of an element $\boldsymbol{v} \in V$ with respect to a basis. For this, we will use the following theorem.

Theorem 17. *Let V be a vector space. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ be linearly independent elements of V. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be numbers. Suppose that we have*

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n = y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2 + \cdots + y_n\boldsymbol{v}_n \tag{8}$$

Then, $x_i = y_i$ for all $i = 1, \ldots, n$.

Proof. It is quite simple to see form (Eq. 8) that $(x_1 - y_1)\,\boldsymbol{v}_1 + (x_2 - y_2)\,\boldsymbol{v}_2 + \cdots + (x_n - y_n)\,\boldsymbol{v}_n = O$. Thus, we have that, given that $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ are linearly independent, $x_i - y_i = 0$ for all $i = 1, 2, ..., n$. □

Let V be a vector space, and let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ be a basis of V. It is possible to represent all $\boldsymbol{v} \in V$ by an n-tuple of numbers relative to this basis as follows

$$\boldsymbol{v} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n \tag{9}$$

Thus, this n-tuple is uniquely determined by $\boldsymbol{v}$. We will call $(x_1, x_2, \ldots, x_n)$ as the coordinates of $\boldsymbol{v}$ with respect to the basis, and we call x_i as the i^{th} coordinate. It is more, the coordinates with respect to the usual basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_n\}$ of K^n are the coordinates of the $n-$tuple $X = (x_1, x_2, \ldots, x_n)$ which is the **coordinate vector** of $\boldsymbol{v}$ with respect to the basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$.

Example 18. Let V be the vector space of functions generated by the three functions e^t, e^{2t}, e^{3t}. Then, the coordinates of the function

$$2e^t + e^{2t} + 10e^{3t} \tag{10}$$

are the 3-tuple $(2, 1, 10)$ with respect to the basis $\left\{e^t, e^{2t}, e^{3t}\right\}$.

3.2 Properties of a Basis

Now, we describe a series of important properties coming from the concept of a basis.

Theorem 19. *(Limit in the size of the basis) Let V be a vector space over a field K with a basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_m\}$. Let $\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_n$ be elements of V, and assume that $n > m$. Then $\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_n$ are linearly dependent.*

Proof. Assume that $\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_n$ are linearly independent. Since $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_m\}$ is a basis V, there exists scalars $a_1, a_2, ..., a_m \in K$ such that

$$\boldsymbol{w}_1 = \sum_{i=1}^{m} a_{ij}\boldsymbol{v}_i$$

By assumption, we know that $\boldsymbol{w}_1 \neq O$, and hence some $a_i \neq 0$. We may assume without loss of generality that $a_1 \neq 0$. We can then solve for $\boldsymbol{v}_1$, and get

$$
\begin{aligned}
a_1\boldsymbol{v}_1 &= \boldsymbol{w}_1 - a_2\boldsymbol{v}_2 - \cdots - a_m\boldsymbol{v}_m, \\
\boldsymbol{v}_1 &= a_1^{-1}\boldsymbol{w}_1 - a_1^{-1}a_2\boldsymbol{v}_2 - \cdots - a_1^{-1}a_m\boldsymbol{v}_m.
\end{aligned}
$$

The subspace of V generated by $\boldsymbol{w}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_m$ contains $\boldsymbol{v}_1$, and hence must be all of V since $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_m$ generate V. Now, we continue our procedure stepwise, and to replace successively $\boldsymbol{v}_2, \boldsymbol{v}_3, ...$ by $\boldsymbol{w}_2, \boldsymbol{w}_3, ...$ until all the elements $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_m$ are exhausted, and $\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_m$ generate V. Now, assume by induction that there is an integer r with $1 \leq r < m$ such that, $\boldsymbol{w}_1, ..., \boldsymbol{w}_r, \boldsymbol{v}_{r+1}, ..., \boldsymbol{v}_m$ generate V. Then, there are elements $b_1, ..., b_r, c_{r+1}, ..., c_m$ in K such that

$$\boldsymbol{w}_{r+1} = b_1\boldsymbol{w}_1 + \cdots + b_r\boldsymbol{w}_r + c_{r+1}\boldsymbol{v}_{r+1} + \cdots + c_m\boldsymbol{v}_m.$$

We cannot have $c_j = 0$ for $j = r+1, ..., m$, if not we get a relation of linear dependence between $\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_r, \boldsymbol{w}_{r+1}$. Then, without loss of generality we can say $c_{r+1} \neq 0$. Thus, we get

$$c_{r+1}\boldsymbol{v}_{r+1} = \boldsymbol{w}_{r+1} - b_1\boldsymbol{w}_1 - \cdots - b_r\boldsymbol{w}_r - c_{r+2}\boldsymbol{v}_{r+2} - \cdots - c_m\boldsymbol{v}_m.$$

Dividing by c_{r+1}, we conclude that $\boldsymbol{v}_{r+1}$ is in the subspace generated by

$$\boldsymbol{w}_1, ..., \boldsymbol{w}_{r+1}, \boldsymbol{v}_{r+2}, ..., \boldsymbol{v}_m.$$

By the induction assumption, we have that

$$\boldsymbol{w}_1, ..., \boldsymbol{w}_{r+1}, \boldsymbol{v}_{r+2}, ..., \boldsymbol{v}_m$$

generate V. Thus by induction, we have proved that $\boldsymbol{w}_1, ..., \boldsymbol{w}_m$ generate V. If $n > m$, then exist elements $d_1, d_2, ..., d_m \in K$ such that

$$\boldsymbol{w}_n = d_1\boldsymbol{w}_1 + \cdots + d_m\boldsymbol{w}_m$$

Thus, $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$ are linearly dependent.

□

Example 20. The vector space $\mathbb{C}^n$ has dimension n over $\mathbb{C}$, the vector space. More generally for any field K, the vector space K^n has dimension n over K with basis

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, ..., \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \tag{11}$$

Example 21. Let V be a vector space. A subspace of dimension 1 is called a line in V. A subspace of dimension 2 is called a plane in V.

Now, we will define the dimension of a vector space V over K which will be denoted by $\dim_K V$, or simply $\dim V$. Then, a vector space with a basis consisting of a finite number of elements, or the zero vector space, is called a **finite dimensional**. Therefore, the $\dim V$ is equal to the number of elements in the finite basis of V. Additionally, vector spaces that do not have a finite basis are called **infinite dimensional**, and although it is possible to define the concept of infinite basis, we will leave for another time.

Thus, the classic question is When a set of vectors is a basis? For this, we have the concept of a **maximal set of linearly independent elements** $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ of V if given any element $\boldsymbol{w} \in V$, the elements $\boldsymbol{w}, \boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ are linearly dependent.

Theorem 22. *Let V be a vector space, and $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ a maximal set of linearly independent elements of V. Then, $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ is a basis of V.*

Proof. We must show $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ generates V. Let $\boldsymbol{w}$ be an element of V. The elements $\boldsymbol{w}, \boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ must be linearly dependent by hypothesis. Therefore there exist scalars $x_0, x_1, ..., x_n$ not all 0 such that

$$x_0\boldsymbol{w} + x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n = O.$$

It is clear that $x_0 \neq 0$ because if that was the case $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ will have a relation of linear dependence among them. Therefore, we have

$$\boldsymbol{w} = -\frac{x_1}{x_0}\boldsymbol{v}_1 - \cdots - \frac{x_n}{x_0}\boldsymbol{v}_n.$$

This is true for all $\boldsymbol{w}$ i.e. a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ and hence a basis. □

Now, we can easily recognize a basis because the following theorem.

Theorem 23. *Let V be a vector space of dimension n, and let $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ be linearly independent elements of V. Then, $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ constitutes a basis of V.*

Proof. According the proof in Theorem 19, $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ is a maximal set of linearly independent elements of V. Therefore, it is a basis by Theorem 22. □

Corollary 24. *Let V be a vector space and let W be a subspace. If* $\dim W = \dim V$ *then $V = W$.*

Proof. A basis for W must also be a basis for V by Theorem 23. □

Corollary 25. *Let V be a vector space of dimension n. Let r be a positive integer with $r < n$, and let $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_r$ be linearly independent elements of V. Then one can find elements $\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, ..., \boldsymbol{v}_n$ such that $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ is a basis of V.*

Proof. Since $r < n$ we know that $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_r\}$ cannot form a basis of V, and it is not a maximal set of linearly independent elements of V. In particular, we can find $\boldsymbol{v}_{r+1} \in V$ such that $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_r, \boldsymbol{v}_{r+1}$ are linearly independent. If $r + 1 < n$, we can repeat the argument. We can this proceed stepwise by induction until we obtain n linearly dependent elements $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n\}$ Then, by Theorem 23 they are a basis. □

Theorem 26. *Let V be a vector space having a basis consisting of n elements. Let W be a subspace which does not consist of O alone. Then W has a basis, and the dimension of W is $\leq n$.*

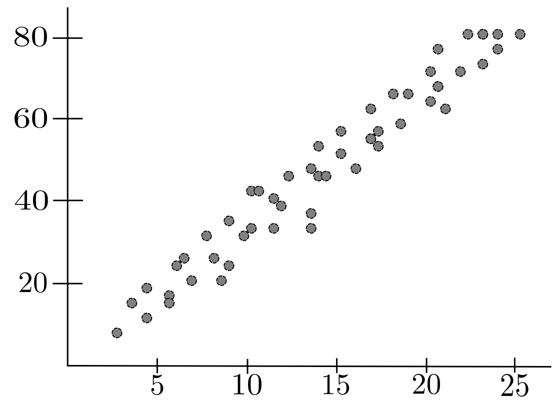

Figure 4: The Points for the Fitting.

Proof. Let $\boldsymbol{w}_1$ be anon-zero element of W. if $\{\boldsymbol{w}_1\}$ is not a maximal set of linearly independent elements of W, we can find an element $\boldsymbol{w}_2$ of W such that $\boldsymbol{w}_1, \boldsymbol{w}_2$ are linerly independent. Thus, we can proceed that way, one element at a time, therefore there must be an integer $m \leq n$ such that $\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_m$ are linear independent and is a maximal set by Theorem 19. Now, using Theorem 22, we can conclude that $\{\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_m\}$ is a basis for W. □

4 Applications in Machine Learning

A classic application of the previous concepts is the vector of features for each sample in the dataset.

Definition 27. A **feature vector** is a n-dimensional vector of numerical features that represent an object.

This allows to use linear algebra to represent basic classification algorithms because the tuples $\{(\boldsymbol{x}, y) \,|\boldsymbol{x} \in K^n, y \in K\}$ can be easily used to design specific algorithms. For example, given the least squared error [5, 6], we need to fit a series of points (Fig. 4) against an specific function. Then, the general problem is given a set of functions $f_1, f_2, ..., f_K$ find values of coefficients $a_1, a_2, ..., a_k$ such that the linear combination

$$y = a_1 f_1(x) + \cdots + a_K f_K(x) \tag{12}$$

For example, Gauss seemed to have used an earlier version of the Leas Squared Error [6] and the following approximation to short arcs.

$$a = z + y \sin^2 L, \tag{13}$$

in order to calculate the meter equal to the one 10,000,000th part of the meridian quadrant (Fig. 5). For more on it, please take a look at the article [6].

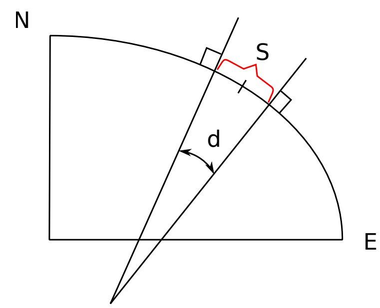

Figure 5: A Meridian Quadrant form the equator (E) to the north pole (N), showing an arc segment of d degrees and length S modules, center at latitude L.

Now, going back to the Least Squared Error, we have that given the datasets $\{(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_N, y_N)\}$, it is possible to define the sample mean as

$$\overline{\boldsymbol{x}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i. \tag{14}$$

Thus, looking at the following data sets in the points $\boldsymbol{x}$ when the sample space V has $\dim V = 1$.

$$\{10, 20, 30, 40, 50\} \text{ and } \{30, 30, 30, 30, 30\}$$

We notice that both have the same sample mean, but they have huge different variances, which is denoted as:

$$\sigma_{\boldsymbol{x}}^2 = \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}).$$

Therefore, we can use this as a tool to measure how much a dataset fluctuates around the mean. Furthermore, by using this idea, we can quantify the meaning of "best fit." For this we assume that our data is coming from a linear equation $y = ax + b$, then $y - (ax + b) \approx 0$. Thus, given the observations $\{(x_1, y_1), ..., (x_N, y_N)\}$, we get the following errors:

$$\{y_1 - (ax_1 + b), ..., y_N - (ax_N + b)\}.$$

Then, the mean should be really small (If it is a good fit), and the variance measures how good fit we have:

$$\sigma_{y-(ax+b)}^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - (ax_i + b))^2. \tag{15}$$

Here, large errors are given a higher weight than smaller errors (due to the squaring). Thus, this procedure favors many medium sized errors over a few large errors. It would be possible to do something similar if we used the absolute value $|y - (ax + b)|$, but this is not differentiable.

Finally, we can define the following error $E_i(a, b) = y - (ax + b)$ making possible to obtain the quadratic accumulated error

$$E(a, b) = \sum_{i=1}^{N} E_i(a, b) = \sum_{i=1}^{N} (y_i - (ax_i + b)) \tag{16}$$

Therefore, the goal is to minimize the previous equation by finding the values a and b. We can do that by differentiating partially in the following way

$$\frac{\partial E}{\partial a} = 0,$$
$$\frac{\partial E}{\partial b} = 0.$$

Note we do not have to worry about boundary point because as $|a|$ and $|b|$ become large, the fill will get worse and worse. Therefore, we do not need to check on boundary.

We get finally,

$$\frac{\partial E}{\partial a} = \sum_{i=1}^{N} 2(y_i - (ax_i + b)) \cdot (-x_i),$$
$$\frac{\partial E}{\partial b} = \sum_{i=1}^{N} 2(y_i - (ax_i + b)) \cdot (-1).$$

Then, we obtain the following

$$\sum_{i=1}^{N} (y_i - (ax_i + b)) \cdot x_i = 0,$$
$$\sum_{i=1}^{N} (y_i - (ax_i + b)) = 0.$$

Finally, by rewriting the previous equations

$$a \sum_{i=1}^{N} x_i^2 + b \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i y_i$$
$$a \sum_{i=1}^{N} x_i + bN = \sum_{i=1}^{N} y_i$$

Then, it is possible to rewrite this in therms of matrices and vectors

$$\begin{pmatrix} \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & N \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i \end{pmatrix} \tag{17}$$

Do Remember this? Yes, the linear representation $A\boldsymbol{x} = \boldsymbol{b}$. Now, we will continue with this in the next class to obtain the optimal solution.

References

[1] Michael Theodore Eismann and SPIE (Society)". *Hyperspectral remote sensing*. Bellingham, Wash. SPIE Press, 2012.

[2] K. Hoffman and R.A. Kunze. *Linear algebra*. Prentice-Hall mathematics series. Prentice-Hall, 1971.

[3] I. Kleiner. *A History of Abstract Algebra*. Birkhäuser Boston, 2007.

[4] S. Lang. *Linear Algebra*. Springer Undergraduate Texts in Mathematics and Technology. Springer, 1987.

[5] S.J. Miller. The method of least squares. Mathematics Department Brown University, 2006.

[6] Stephen M. Stigler. Gauss and the invention of least squares. *The Annals of Statistics*, 9(3):465–474, 1981.