Machine Learning

Linear algebra revisited

Emanuele Rodolà
rodola@di.uniroma1.it

SAPIENZA
UNIVERSITÀ DI ROMA

2nd semester a.y. 2023/2024 · March 5, 2024

Linear algebra is the study of linear maps on finite dimensional vector spaces

Linear algebra is about matrices as much as astronomy is about telescopes

Vector space

The motivation for the definition of a vector space comes from the classical properties of addition and scalar multiplication.

Vector space

The motivation for the definition of a vector space comes from the classical properties of addition and scalar multiplication.

A vector space V is a set along with addition and scalar multiplication such that:

- **commutativity**: $u + v = v + u$ for all $u, v \in V$; further, $u + v \in V$

Vector space

The motivation for the definition of a vector space comes from the classical properties of addition and scalar multiplication.

A vector space V is a set along with addition and scalar multiplication such that:

- **commutativity**: $u + v = v + u$ for all $u, v \in V$; further, $u + v \in V$
- **associativity**: $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbb{R}$; further, $av \in V$

Vector space

The motivation for the definition of a vector space comes from the classical properties of addition and scalar multiplication.

A vector space V is a set along with addition and scalar multiplication such that:

- **commutativity**: $u + v = v + u$ for all $u, v \in V$; further, $u + v \in V$
- **associativity**: $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbb{R}$; further, $av \in V$

"what happens in Vegas, stays in Vegas"

Vector space

The motivation for the definition of a vector space comes from the classical properties of addition and scalar multiplication.

A vector space V is a set along with addition and scalar multiplication such that:

- **commutativity**: $u + v = v + u$ for all $u, v \in V$; further, $u + v \in V$
- **associativity**: $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbb{R}$; further, $av \in V$

"what happens in Vegas, stays in Vegas"

- **additive identity**: there exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$

Vector space

The motivation for the definition of a vector space comes from the classical properties of addition and scalar multiplication.

A vector space V is a set along with addition and scalar multiplication such that:

- **commutativity**: $u + v = v + u$ for all $u, v \in V$; further, $u + v \in V$
- **associativity**: $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbb{R}$; further, $av \in V$

"what happens in Vegas, stays in Vegas"

- **additive identity**: there exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$
- **additive inverse**: for every $v \in V$, there exists $w \in V$ such that $v + w = 0$

Vector space

The motivation for the definition of a vector space comes from the classical properties of addition and scalar multiplication.

A vector space V is a set along with addition and scalar multiplication such that:

- **commutativity**: $u + v = v + u$ for all $u, v \in V$; further, $u + v \in V$
- **associativity**: $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbb{R}$; further, $av \in V$

"what happens in Vegas, stays in Vegas"

- **additive identity**: there exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$
- **additive inverse**: for every $v \in V$, there exists $w \in V$ such that $v + w = 0$
- **multiplicative identity**: $1v = v$ for all $v \in V$

Vector space

The motivation for the definition of a vector space comes from the classical properties of addition and scalar multiplication.

A vector space V is a set along with addition and scalar multiplication such that:

- **commutativity**: $u + v = v + u$ for all $u, v \in V$; further, $u + v \in V$
- **associativity**: $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbb{R}$; further, $av \in V$

"what happens in Vegas, stays in Vegas"

- **additive identity**: there exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$
- **additive inverse**: for every $v \in V$, there exists $w \in V$ such that $v + w = 0$
- **multiplicative identity**: $1v = v$ for all $v \in V$
- **distributive properties**: $a(u + v) = au + av$ and $(a + b)v = av + bv$ for all $a, b \in \mathbb{R}$ and all $u, v \in V$

Example: Lists of numbers

$\mathbb{R}^n$ is defined to be the set of all n-long sequences of numbers in $\mathbb{R}$:

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_j \in \mathbb{R} \text{ for } j = 1, 2, \ldots, n\}$$

Example: Lists of numbers

$\mathbb{R}^n$ is defined to be the set of all n-long sequences of numbers in $\mathbb{R}$:

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_j \in \mathbb{R} \text{ for } j = 1, 2, \ldots, n\}$$

Addition and multiplication are defined as expected:

$$\begin{aligned}(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \\ \lambda(x_1, x_2, \ldots, x_n) &= (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)\end{aligned}$$

Example: Lists of numbers

$\mathbb{R}^n$ is defined to be the set of all n-long sequences of numbers in $\mathbb{R}$:

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_j \in \mathbb{R} \text{ for } j = 1, 2, \ldots, n\}$$

Addition and multiplication are defined as expected:

$$\begin{aligned}(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \\ \lambda(x_1, x_2, \ldots, x_n) &= (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)\end{aligned}$$

While the additive identity can be defined as:

$$0 = (0, \ldots, 0)$$

Example: Lists of numbers

$\mathbb{R}^n$ is defined to be the set of all n-long sequences of numbers in $\mathbb{R}$:

$$\mathbb{R}^n = \{(x_1, \dots, x_n) : x_j \in \mathbb{R} \text{ for } j = 1, 2, \dots, n\}$$

Addition and multiplication are defined as expected:

$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$
$$\lambda(x_1, x_2, \dots, x_n) = (\lambda x_1, \lambda x_2, \dots, \lambda x_n)$$

While the additive identity can be defined as:

$$0 = (0, \dots, 0)$$

With these definitions, $\mathbb{R}^n$ is a vector space

Example: Functions

Consider the set of all functions $f : [0,1] \to \mathbb{R}$ with the standard definitions for sum and scalar product:

$$(f+g)(x) = f(x) + g(x)$$
$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in [0,1]$ and $\lambda \in \mathbb{R}$

Example: Functions

Consider the set of all functions $f : [0, 1] \to \mathbb{R}$ with the standard definitions for sum and scalar product:

$$(f + g)(x) = f(x) + g(x)$$
$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in [0, 1]$ and $\lambda \in \mathbb{R}$

and with additive identity and inverse defined as:

$$0(x) = 0$$
$$(-f)(x) = -f(x)$$

for all $x \in [0, 1]$

Example: Functions

Consider the set of all functions $f : [0,1] \to \mathbb{R}$ with the standard definitions for sum and scalar product:

$$(f+g)(x) = f(x) + g(x)$$
$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in [0,1]$ and $\lambda \in \mathbb{R}$

and with additive identity and inverse defined as:

$$0(x) = 0$$
$$(-f)(x) = -f(x)$$

for all $x \in [0,1]$

The above forms a vector space. In fact, any set of functions $f : S \to \mathbb{R}$ with $S \neq \emptyset$ (Q: why?) and the definitions above forms a vector space.

Elements of a vector space (called vectors)
are not necessarily lists

A vector space is an abstract entity whose elements
might be lists, functions, or weird objects

Example: Curved surfaces

Do surfaces form a vector space?

Example: Curved surfaces

Do surfaces form a vector space? Not really – if you sum the coordinates of two points, you may get a third point that is not on the surface.

Example: Curved surfaces

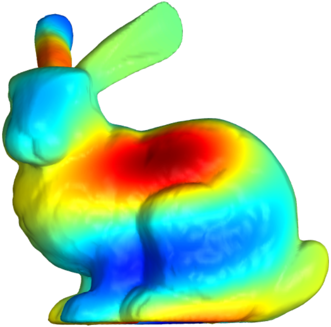
Do surfaces form a vector space? Not really – if you sum the coordinates of two points, you may get a third point that is not on the surface.

Surfaces can be studied using differential geometry.

Example: Curved surfaces

Do surfaces form a vector space? Not really – if you sum the coordinates of two points, you may get a third point that is not on the surface.

Surfaces can be studied using differential geometry.



We can still use linear algebra to study functions on surfaces

Bases

A basis of V is a collection of vectors in V that is linearly independent and spans V.

Bases

A basis of V is a collection of vectors in V that is linearly independent and spans V.

- $\mathrm{span}(v_1, \ldots, v_n) = \{a_1 v_1 + \cdots + a_n v_n : a_1, \ldots, a_n \in \mathbb{R}\}$

Bases

A basis of V is a collection of vectors in V that is linearly independent and spans V.

- $\text{span}(v_1, \ldots, v_n) = \{a_1 v_1 + \cdots + a_n v_n : a_1, \ldots, a_n \in \mathbb{R}\}$
- $v_1, \ldots, v_n \in V$ are linearly independent if and only if each $v \in \text{span}(v_1, \ldots, v_n)$ has only one representation as a linear combination of $v_1, \ldots, v_n$

Bases

A basis of V is a collection of vectors in V that is linearly independent and spans V.

- $\text{span}(v_1, \ldots, v_n) = \{a_1 v_1 + \cdots + a_n v_n : a_1, \ldots, a_n \in \mathbb{R}\}$
- $v_1, \ldots, v_n \in V$ are linearly independent if and only if each $v \in \text{span}(v_1, \ldots, v_n)$ has only one representation as a linear combination of $v_1, \ldots, v_n$

So every vector $v \in V$ can be expressed uniquely as a linear combination

$$v = \sum_{i=1}^{n} \alpha_i v_i$$

You can think of a basis as the minimal set of vectors that generates the entire space

Example: Bases

- $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ is a basis of $\mathbb{R}^n$ called the standard basis; its vectors are called the indicator vectors.

  In deep learning, also called one-hot representation.

Example: Bases

- $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ is a basis of $\mathbb{R}^n$ called the standard basis; its vectors are called the indicator vectors.

  In deep learning, also called one-hot representation.

- $(1, 2), (3, 5.07)$ is a basis of $\mathbb{R}^2$

Example: Bases

- $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ is a basis of $\mathbb{R}^n$ called the standard basis; its vectors are called the indicator vectors.

  In deep learning, also called one-hot representation.

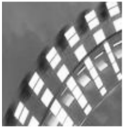
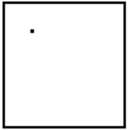
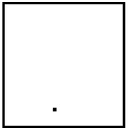
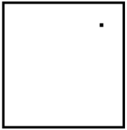
- $(1, 2), (3, 5.07)$ is a basis of $\mathbb{R}^2$

- $$f_1(x) = \begin{cases} 1 & \text{if } x = x_1 \\ 0 & \text{else} \end{cases}$$
  $$f_2(x) = \begin{cases} 1 & \text{if } x = x_2 \\ 0 & \text{else} \end{cases}$$
  $$\vdots$$

  is the standard basis for the set of functions $f : \mathbb{R} \to \mathbb{R}$; the basis vectors are also called indicator functions

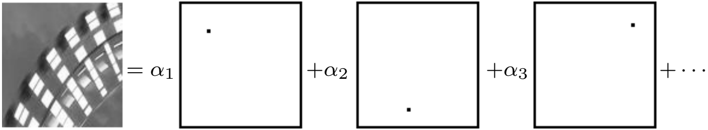
Examples

An image expressed in the standard basis:

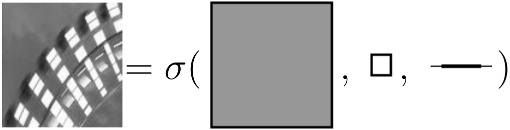
 $= \alpha_1$  $+\alpha_2$  $+\alpha_3$  $+\cdots$

Examples

An image expressed in the standard basis:



The same image, expressed in terms of a nonlinear map σ:



The image is **not** in the span of the three features.

Dimension

A vector space may have different bases; any two bases have the same number of vectors

Dimension

A vector space may have different bases; any two bases have the same number of vectors

The dimension of a vector space is the number of basis vectors

Dimension

A vector space may have different bases; any two bases have the same number of vectors

The dimension of a vector space is the number of basis vectors

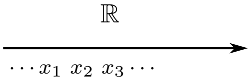
Note: Even though function spaces are not necessarily finite dimensional (Q: why?), with digital data they usually are, since we deal with finite discrete domains (images, graphs, text, etc.)

Dimension

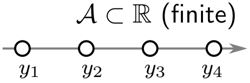
A vector space may have different bases; any two bases have the same number of vectors

The dimension of a vector space is the number of basis vectors

Note: Even though function spaces are not necessarily finite dimensional (Q: why?), with digital data they usually are, since we deal with finite discrete domains (images, graphs, text, etc.)



$f : \mathbb{R} \to \mathbb{R}$
infinite dimensional
(functional analysis)



$f : \mathcal{A} \to \mathbb{R}$
finite dimensional
(linear algebra)

Linear maps

A linear map from V to W is a function $T : V \to W$ with the properties:

- **additivity**: $T(u + v) = Tu + Tv$ for all $u, v \in V$
- **homogeneity**: $T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbb{R}$ and all $v \in V$

Linear maps

A linear map from V to W is a function $T : V \to W$ with the properties:

- **additivity**: $T(u + v) = Tu + Tv$ for all $u, v \in V$
- **homogeneity**: $T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbb{R}$ and all $v \in V$

Examples:

- identity $I : V \to V$, defined as $Iv = v$

Linear maps

A linear map from V to W is a function $T : V \to W$ with the properties:

- **additivity**: $T(u + v) = Tu + Tv$ for all $u, v \in V$
- **homogeneity**: $T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbb{R}$ and all $v \in V$

Examples:

- identity $I : V \to V$, defined as $Iv = v$
- differentiation $D : \mathcal{F}(\mathbb{R}) \to \mathcal{F}(\mathbb{R})$, defined as $Df = f'$

Linear maps

A linear map from V to W is a function $T : V \to W$ with the properties:

- **additivity**: $T(u+v) = Tu + Tv$ for all $u, v \in V$
- **homogeneity**: $T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbb{R}$ and all $v \in V$

Examples:

- identity $I : V \to V$, defined as $Iv = v$
- differentiation $D : \mathcal{F}(\mathbb{R}) \to \mathcal{F}(\mathbb{R})$, defined as $Df = f'$
- integration $T : \mathcal{F}(\mathbb{R}) \to \mathbb{R}$, defined as $Tf = \int_0^1 f(x)dx$

Linear maps

A linear map from V to W is a function $T : V \to W$ with the properties:

- **additivity**: $T(u + v) = Tu + Tv$ for all $u, v \in V$
- **homogeneity**: $T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbb{R}$ and all $v \in V$

Examples:

- identity $I : V \to V$, defined as $Iv = v$
- differentiation $D : \mathcal{F}(\mathbb{R}) \to \mathcal{F}(\mathbb{R})$, defined as $Df = f'$
- integration $T : \mathcal{F}(\mathbb{R}) \to \mathbb{R}$, defined as $Tf = \int_0^1 f(x)dx$
- a map $T : \mathbb{R}^3 \to \mathbb{R}^2$, defined as

$$T(x, y, z) = (2x - y + 3z, 7x + 5y - 6z)$$

Linear maps

A linear map from V to W is a function $T : V \to W$ with the properties:

- **additivity**: $T(u+v) = Tu + Tv$ for all $u, v \in V$
- **homogeneity**: $T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbb{R}$ and all $v \in V$

Examples:

- identity $I : V \to V$, defined as $Iv = v$
- differentiation $D : \mathcal{F}(\mathbb{R}) \to \mathcal{F}(\mathbb{R})$, defined as $Df = f'$
- integration $T : \mathcal{F}(\mathbb{R}) \to \mathbb{R}$, defined as $Tf = \int_0^1 f(x)dx$
- a map $T : \mathbb{R}^3 \to \mathbb{R}^2$, defined as

$$T(x, y, z) = (2x - y + 3z, 7x + 5y - 6z)$$

- a map $T : \mathbb{R}^n \to \mathbb{R}^m$, defined as

$$T(x_1, \ldots, x_n) = (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \ldots, A_{m,1}x_1 + \cdots + A_{m,n}x_n)$$

Examples: Are these linear?

Equation of a line:

$$y = ax + b$$

Examples: Are these linear?

Equation of a line:

$$y = ax + b$$

In deep learning, the term b is also called a bias.

Examples: Are these linear?

Equation of a line:

$$y = ax + b$$

In deep learning, the term b is also called a bias.

This other equation:

$$y = z\sin(x) + z^2\sin(x)$$

Examples: Are these linear?

Equation of a line:

$$y = ax + b$$

In deep learning, the term b is also called a bias.

This other equation:

$$y = z\sin(x) + z^2\sin(x)$$

...wrt which variable?

Examples: Are these linear?

Equation of a line:

$$y = ax + b$$

In deep learning, the term b is also called a bias.

This other equation:

$$y = z\sin(x) + z^2\sin(x)$$

...wrt which variable?

Reflection operation on an image:

$$T : \mathbb{R}^2 \to \mathbb{R}^2\,, \quad T(x, y) = (-x, y)$$

$\Rightarrow$

Linear maps as a vector space

Linear maps $T : V \to W$ form a vector space, with addition and multiplication (Q: what is the additive identity?) defined as:

$$(S + T)(v) = Sv + Tv$$
$$(\lambda T)(v) = \lambda(Tv)$$

Linear maps as a vector space

Linear maps $T : V \to W$ form a vector space, with addition and multiplication (Q: what is the additive identity?) defined as:

$$(S + T)(v) = Sv + Tv$$
$$(\lambda T)(v) = \lambda(Tv)$$

We also have a useful definition of product between linear maps.

Linear maps as a vector space

Linear maps $T : V \to W$ form a vector space, with addition and multiplication (Q: what is the additive identity?) defined as:

$$(S + T)(v) = Sv + Tv$$
$$(\lambda T)(v) = \lambda(Tv)$$

We also have a useful definition of product between linear maps.

If $T : U \to V$ and $S : V \to W$, their product $ST : U \to W$ is defined by

$$(ST)(u) = S(Tu)$$

In other words, ST is just the usual composition $S \circ T$ of two functions

Algebraic properties of products of linear maps

- **associativity**: $(T_1T_2)T_3 = T_1(T_2T_3)$
- **identity**: $TI = IT = T$
- **distributive properties**: $(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$

Algebraic properties of products of linear maps

- **associativity**: $(T_1T_2)T_3 = T_1(T_2T_3)$
- **identity**: $TI = IT = T$
- **distributive properties**: $(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$

Keep in mind that composition of linear maps is not commutative, i.e.

$$ST \neq TS$$

in general (although there are special cases)

Example: Take $Sf = f'$ and $(Tf)(x) = x^2 f(x)$

Matrices

Consider a linear map $T : V \to W$, a basis $v_1, \ldots, v_n \in V$ and a basis $w_1, \ldots, w_m \in W$.

Matrices

Consider a linear map $T : V \to W$, a basis $v_1, \ldots, v_n \in V$ and a basis $w_1, \ldots, w_m \in W$.

The matrix of T in these bases is the $m \times n$ array of values in $\mathbb{R}$

$$\mathbf{T} = \begin{pmatrix} T_{1,1} & \cdots & T_{1,n} \\ \vdots & & \vdots \\ T_{m,1} & \cdots & T_{m,n} \end{pmatrix}$$

whose entries $T_{i,j}$ are defined by

$$Tv_j = T_{1,j} w_1 + \cdots + T_{m,j} w_m$$

Matrices

Consider a linear map $T : V \to W$, a basis $v_1, \ldots, v_n \in V$ and a basis $w_1, \ldots, w_m \in W$.

The matrix of T in these bases is the $m \times n$ array of values in $\mathbb{R}$

$$\mathbf{T} = \begin{pmatrix} T_{1,1} & \cdots & T_{1,n} \\ \vdots & & \vdots \\ T_{m,1} & \cdots & T_{m,n} \end{pmatrix}$$

whose entries $T_{i,j}$ are defined by

$$Tv_j = T_{1,j} w_1 + \cdots + T_{m,j} w_m$$

Hence each column of $\mathbf{T}$ contains the linear combination coefficients for the image via T of a basis vector from V

Matrices

Consider a linear map $T : V \to W$, a basis $v_1, \ldots, v_n \in V$ and a basis $w_1, \ldots, w_m \in W$.

The matrix of T in these bases is the $m \times n$ array of values in $\mathbb{R}$

$$\mathbf{T} = \begin{pmatrix} T_{1,1} & \cdots & T_{1,n} \\ \vdots & & \vdots \\ T_{m,1} & \cdots & T_{m,n} \end{pmatrix}$$

whose entries $T_{i,j}$ are defined by

$$Tv_j = T_{1,j} w_1 + \cdots + T_{m,j} w_m$$

In other words, the matrix encodes how basis vectors are mapped, and this is enough to map all other vectors in their span, since:

$$Tv = T(\sum_j \alpha_j v_j) = \sum_j T(\alpha_j v_j) = \sum_j \alpha_j Tv_j$$

The matrix is a representation for a linear map, and it depends on the choice of bases

Matrix of a vector

Suppose $v \in V$ is an arbitrary vector, while $v_1, \ldots, v_n$ is a basis of V. The matrix of v wrt this basis is the $n \times 1$ matrix:

$$\mathbf{v} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

so that

$$v = c_1 v_1 + \cdots + c_n v_n$$

Once again, we see that the matrix depends on the choice of basis for V

Product of "map matrix" and "vector matrix"

$$\underbrace{\begin{pmatrix} T_{1,1} & \cdots & T_{1,n} \\ \vdots & & \vdots \\ T_{m,1} & \cdots & T_{m,n} \end{pmatrix}}_{\mathbf{T}} \underbrace{\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}}_{\mathbf{c}} = \sum_{j=1}^{n} c_j \underbrace{\begin{pmatrix} T_{1,j} \\ \vdots \\ T_{m,j} \end{pmatrix}}_{\mathrm{Tv_j\ \ wrt\ \ (w_1,\ldots,w_m)}}$$

Because recall that, for bases $v_1, \ldots, v_n \in V$ and $w_1, \ldots, w_m \in W$:

$$Tv_j = T_{1,j} w_1 + \cdots + T_{m,j} w_m$$

Product of "map matrix" and "vector matrix"

$$\underbrace{\begin{pmatrix} T_{1,1} & \cdots & T_{1,n} \\ \vdots & & \vdots \\ T_{m,1} & \cdots & T_{m,n} \end{pmatrix}}_{\mathbf{T}} \underbrace{\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}}_{\mathbf{c}} = \sum_{j=1}^{n} c_j \underbrace{\begin{pmatrix} T_{1,j} \\ \vdots \\ T_{m,j} \end{pmatrix}}_{\mathrm{Tv_j \ wrt \ (w_1,\ldots,w_m)}}$$

Because recall that, for bases $v_1, \ldots, v_n \in V$ and $w_1, \ldots, w_m \in W$:

$$Tv_j = T_{1,j} w_1 + \cdots + T_{m,j} w_m$$

We see then that vector $c = \sum_j c_j v_j$ is mapped to $Tc = \sum_j c_j T v_j$.

In other words, matrix product is behaving as expected.

Suggested reading

Sections 1.A – 3.D of the textbook:

S. Axler, "Linear algebra done right – 3rd edition". Springer, 2015