Faculty of Information Engineering, Informatics and Statistics
Department of Computer Science

Deep Learning and Applied AI

Author:
Simone Lidonnici

25 february 2026

Contents

1

Linear algebra

1.1 Vector spaces

A vector space V is a set with an addition and a scalar multiplication that respect some properties:

- **Commutativity**: $u + v = v + u \; \forall u, v \in V$ and also $u + v \in V$
- **Associativity:** $(u + v) + w = u + (v + w) \; \forall u, v, w \in V$ and $(ab)v = a(bv) \; \forall a, b \in \mathbb{R}$ and $av \in V$
- **Additive identity**: $\exists 0 | v + 0 = v \; \forall v \in V$
- **Additive inverse:** $\forall v \in V \; \exists w \in V | v + w = 0$
- **Multiplicative identity**: $\exists 1 \in \mathbb{R} | 1v = v \; \forall v \in V$
- **Distributive**: $a(u + v) = au + av \; \forall u, v \in V \; \forall a \in \mathbb{R}$

Example:
The set of all functions $f : [0, 1] \to \mathbb{R}$ with addition and scalar multiplication as:

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

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

And additive identity and inverse as:

$$0(x) = 0$$

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

This is a vector space and all sets of functions $f : S \to \mathbb{R}$ is a vector space.

1.1.1 Base of a vector space

A base of a vector space V is a set of vectors $v_1, \ldots, v_n \in V$ that are linearly indipendent and spans V.
The span of a set of vectors is defined as:

$$\mathrm{Span}(v_1, \ldots, v_n) = \{\alpha_1 v_1 + \cdots + \alpha_n v_n | \alpha_1, \ldots, \alpha_n \in \mathbb{R}\}$$

So, to span V and be linearly indipendent means that each vector $v \in V$ can be represented by only one linear combination of $v_1, \ldots, v_n$:

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

A vector space has infinite bases, all with the same number of vectors and this number is the dimension of the vector space.

1.2 Linear maps

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

- **Additivity**: $T(u + v) = T(u) + T(v)\ \forall u, v \in V$
- **Homogeneity**: $T(\lambda v) = \lambda T(v)\ \forall v \in V\ \forall \lambda \in \mathbb{R}$

Example:
The equation:

$$y = ax + b$$

is not a linear map, because of the constant b, that in deep learning is called a bias.
This equation:

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

is not a linear map in respect to z or x but is a linear map in respect to $\sin(x)$.

1.2.1 Vector space of linear maps

Fixed two vector spaces V, W, the set of all linear map $T : V \to W$ is itself a vector space with addition and scalar multiplication:

$$(S + T)(v) = S(v) + T(v)$$

$$(\lambda T)(v) = \lambda T(v)$$

The additive identity is the linear map that maps all values to the identity of the vector space W.

1.2.2 Product of linear map

Taken two linear maps $T : V \to W$ and $S : W \to Z$, their product is defined as $ST : V \to Z$:

$$(ST)(v) = S(T(v))$$

So, is the composition $S \circ T$. The product has different properties:

- **Associativity**: $(T_1T_2)T_3 = T_1(T_2T_3)$
- **Identity**: $\exists I | TI = IT = T$
- **Distributive**: $(S_1 + S_2)T = S_1T + S_2T$

The product is not commutative in general.

1.3 Matrices

Let $T : V \to W$ be a linear map, $v_1, \dots, v_n \in V$ the base of V and $w_1, \dots, w_m \in W$ the base of W. The map T can be represented by a $m \times n$ matrix of values in $\mathbb{R}$:

$$T = \begin{vmatrix} T_{1,1} & \dots & T_{1,n} \\ \vdots & \ddots & \vdots \\ T_{m,1} & \dots & T_{m,n} \end{vmatrix}$$

The entries of the matrix are defined in such a way that:

$$T(v_j) = T_{1,j}w_1 + \dots + T_{m,j}w_m$$

The matrix encodes how the vectors of the base are mapped, but it also maps all other vector in V:

$$T(v) = T(\sum_{\forall j} \alpha_j v_j) = \sum_{\forall j} T(\alpha_j v_j) = \sum_{\forall j} \alpha_j T(v_j)$$

The matrix representation of T depends on the base chosed.

1.3.1 Vector represented as matrix

In a vector space V with base $v_1, \dots, v_n$ an arbitrary vector $v \in V$ can be defined as a $n \times 1$ matrix:

$$v = \begin{vmatrix} c_1 \\ \vdots \\ c_n \end{vmatrix}$$

In which the coefficient are such that:

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

1.3.2 Matrix manipulation

A matrix is symmetric if:

$$A = A^T \implies A_{i,j} = A_{j,i} \; \forall i, j$$

For a product $A = BC$ the transpose and inverse are:

$$A^T = (BC)^T = C^T B^T$$

$$A^{-1} = (BC)^{-1} = C^{-1} B^{-1}$$

A matrix is orthogonal if:

$$A^{-1} = A^T$$

And when the matrix is orthogonal we have that $A^T A = I$.

1.3.3 Products with matrices

There a different ways in which matrices can be multiplied.
Matrix-Vector product:

$$Xy = \begin{vmatrix} | & \dots & | \\ x_1 & & x_n \\ | & \dots & | \end{vmatrix} \begin{vmatrix} y_1 \\ \vdots \\ y_n \end{vmatrix} = \sum_{j=1}^{n} y_j \begin{vmatrix} | \\ x_j \\ | \end{vmatrix}$$

Vector-Matrix product:

$$y^T X = (X^T y)^T$$

Matrix-Matrix product:

$$XY = \begin{vmatrix} - & x_1^T & - \\ \vdots & & \vdots \\ - & x_m^T & - \end{vmatrix} \begin{vmatrix} | & \dots & | \\ y_1 & & y_n \\ | & \dots & | \end{vmatrix} = \begin{vmatrix} | & \dots & | \\ Xy_1 & & Xy_n \\ | & \dots & | \end{vmatrix} = \begin{vmatrix} - & x_1^T Y & - \\ \vdots & & \vdots \\ - & x_m^T Y & - \end{vmatrix}$$

Vector-Vector product (inner):

$$x^T y = \alpha$$

Vector-Vector product (outer):

$$xy^T = \begin{vmatrix} x_1 \\ \vdots \\ x_m \end{vmatrix} \begin{vmatrix} y_1 & \dots & y_n \end{vmatrix} = \begin{vmatrix} | & \dots & | \\ y_1 x & & y_n x \\ | & \dots & | \end{vmatrix}$$

1.3.4 Matrix parameters

The trace of a matrix is the sum of the diagonal:

$$tr(A) = \sum_{\forall i} a_{i,i}$$

The trace is a linear map, since:

$$tr(A + B) = tr(A) + tr(B)$$

$$tr(\alpha A) = \alpha tr(A)$$

It is also invariant to transpose and cyclic permutations:

$$tr(A) = tr(A^T)$$

$$tr(ABC) = tr(CAB) = tr(BCA)$$

The gradient of the trace of a matrix:

$$\nabla tr(A) = \nabla \sum_{\forall i} a_{i,i}$$

Is computed by partial derivatives:

$$\frac{\partial}{\partial a_{i,j}} \sum_{\forall i} a_{i,i}$$

The result is the identity matrix, because the derivatives are 1 only for an element in the diagonal like $a_{i,i}$.
The Frobenius norm of a matrix is:

$$||X||_F^2 = tr(XX^T) = tr(X^TX)$$

For a vector is much simple:

$$||x||_F^2 = tr(xx^T) = tr(x^Tx) = x^Tx$$

With this we can calculate the distance between two matrices:

$$||A - B||_F^2 = tr((A - B)^T(A - B)) = tr(A^TA) + tr(B^TB) - 2tr(A^TB)$$

The 1 vector (the vector with all 1) can be used to calculate sum of values of a matrix easy. To calculate the sum of values on each row:

$$A1$$

The sum of values of each column:

$$1^TA$$

The sum of all values:

$$1^TA1$$

Also we have that:

$$x^TAx = tr(x^TAx) = tr(xx^TA) = tr(XA)$$

In which $X = xx^T$ and creating this matrix is called lifting.
The permutation matrices are orthogonal matrices that are doubly stochastic and have a single 1 in every column and row:

$$\begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} 1 \\ 2 \\ 3 \end{vmatrix} = \begin{vmatrix} 2 \\ 1 \\ 3 \end{vmatrix}$$

The convex combination of two permutation matrices P, Q is another permutation matrix:

$$\alpha P + (1 - \alpha)Q = D \; \forall \alpha \in [0, 1]$$