Computational Intelligence Lab: Linear Algebra Recap

Cristian Perez Jensen

July 20, 2024

Contents

1 Inner product and norms

Definition 1 (Inner product)**.** An inner product $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ is an operation defined on a vector space $\mathcal{V}$ that satisfies the following properties $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, a, b \in \mathbb{R}$,

- Commutativity,
$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle;$$

- Linearity,
$$\langle \boldsymbol{x}, a\boldsymbol{y} + b\boldsymbol{z} \rangle = a\langle \boldsymbol{x}, \boldsymbol{y} \rangle + b\langle \boldsymbol{x}, \boldsymbol{z} \rangle;$$

- Positive definiteness,
$$\begin{aligned} \boldsymbol{x} \neq \mathbf{0} &\implies \langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0 \\ \boldsymbol{x} = 0 &\iff \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0; \end{aligned}$$

- Bilinearity (follows from commutativity and linearity),
$$\begin{aligned} \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle &= \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle \\ \langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle &= \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle. \end{aligned}$$

Corollary. $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{x} \rangle + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle$.

Corollary. $\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^\top \boldsymbol{y} \rangle$.

The vector space $\mathcal{V}$, along with an inner product, defines an inner vector space. During this course, we will assume that we always work with real vectors in Rn. An example of an inner product is the dot product,[1]

$$\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^\top \boldsymbol{y} \in \mathbb{R}.$$

[1] Usually, this operation is what is meant by the inner product.

Definition 2 (Norm)**.** A norm $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ is a function that can be thought of as a way of measuring the distance from the origin. Norms satisfy the following properties,

- Positive definiteness, $\boldsymbol{x} \neq \mathbf{0} \implies \|\boldsymbol{x}\| > 0$;
- Triangle inequality, $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$;
- Cauchy-Schwarz inequality, $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$.

Corollary. For the Euclidean norm, the following holds,

$$\cos\theta = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|},$$

where θ is the angle between x and y.

Each inner product defines a canonical norm $\|\boldsymbol{x}\| \doteq \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$. For example, the Euclidean norm is defined by the dot product,

$$\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

The p-norm is a generalization of the Euclidean norm,

$$\|\boldsymbol{x}\|_p = \sqrt{\sum_{i=1}^{n} |x_i|^p}.$$

2 *Vector spaces*

The *vector space* $\mathbb{R}^m$ consists of all column vectors with m elements. For a set of vectors $\mathcal{C} = \{\boldsymbol{c}_1, \dots, \boldsymbol{c}_n \mid \boldsymbol{c}_i \in \mathbb{R}^m\}$, we can define a subspace spanned by this set, denoted by $\mathrm{span}(\mathcal{C})$. It is the set of all possible linear combinations of elements of $\mathcal{C}$. If a set of vectors that span a subspace are independent, they are called a *basis*, $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k \in \mathbb{R}^m$. The number of basis vectors defines the *dimensionality* of the subspace.[2]

[2] We know that the amount of basis vectors must be smaller than the amount of vectors that span the subspace, which must be smaller than the dimensionality of the space,

$$k \leq m \leq n.$$

Observation. The following facts hold about subspaces,

- Every subspace contains the zero vector $\mathbf{0}$;
- If $\boldsymbol{x}$ and $\boldsymbol{y}$ are in the subspace, then $\boldsymbol{x} + \boldsymbol{y}$ is also in the subspace;
- If $\boldsymbol{x}$ is in the subspace and $a \in \mathbb{R}$, then $a\boldsymbol{x}$ is also in the subspace.

Definition 3 (Orthogonal subspaces)**.** Subspaces $\mathcal{V}$ and $\mathcal{W}$ are orthogonal when $\boldsymbol{v}^\top \boldsymbol{w} = 0$ for all $\boldsymbol{v} \in \mathcal{V}, \boldsymbol{w} \in \mathcal{W}$.

3 *Matrices*

The rank r of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the dimensionality of its *column space*. It is bounded by

$$r \leq \min\{m, n\}.$$

The matrix is full-rank if $r = \min\{m, n\}$.

A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ defines 4 fundamental subspaces,

- Column space $\subseteq \mathbb{R}^m$ (r dimensional), $\{\boldsymbol{b} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$;
- Null space $\subseteq \mathbb{R}^n$ ($n - r$ dimensional), $\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \mathbf{0}\}$;
- Row space $\subseteq \mathbb{R}^n$ (r dimensional), $\{\boldsymbol{b} \mid \boldsymbol{A}^\top \boldsymbol{x} = \boldsymbol{b}\}$;

- Left null space $\subseteq \mathbb{R}^m$ ($m - r$ dimensional), $\{\boldsymbol{x} \mid \boldsymbol{A}^\top \boldsymbol{x} = \boldsymbol{0}\}$.

The row space $\operatorname{row}(\boldsymbol{A})$ is the *orthogonal complement* of the null space $\operatorname{null}(\boldsymbol{A})$, thus $\operatorname{row}(\boldsymbol{A}) + \operatorname{null}(\boldsymbol{A}) = \mathbb{R}^n$. Similarly, $\operatorname{col}(\boldsymbol{A}) + \operatorname{null}(\boldsymbol{A}^\top) = \mathbb{R}^m$.

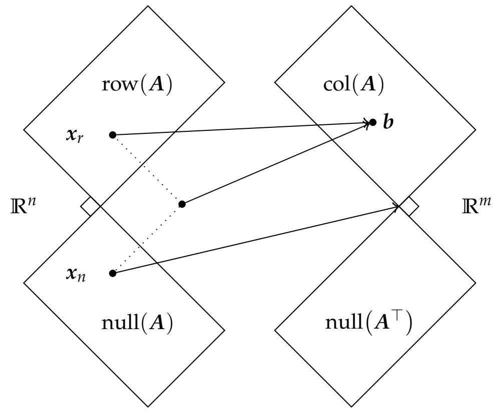

Figure 1. Illustration of the 4 spaces defined by a matrix $\boldsymbol{A}$. It shows the perpendicular spaces. Furthermore, it shows that $\boldsymbol{A}\boldsymbol{x}_r = \boldsymbol{b}$ for some $\boldsymbol{x}_r \in \operatorname{col}(\boldsymbol{A})$. Also, if you add a vector from the null space to the row vector, it still maps to the same $\boldsymbol{b}$, $\boldsymbol{A}(\boldsymbol{x}_r + \boldsymbol{x}_n) = \boldsymbol{A}\boldsymbol{x}_r + \boldsymbol{A}\boldsymbol{x}_n = \boldsymbol{A}\boldsymbol{x}_r = \boldsymbol{b}$.

3.1 *Invertible matrices*

A matrix is only invertible if it is a square full-rank matrix, *i.e.*, $r = m = n$.

Properties. Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a full-rank matrix and $k \in \mathbb{R}$, then

$$\begin{aligned}
\boldsymbol{A}^{-1}\boldsymbol{A} &= \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} \\
(k\boldsymbol{A})^{-1} &= \frac{1}{k}\boldsymbol{A}^{-1} \\
\det\left(\boldsymbol{A}^{-1}\right) &= \frac{1}{\det(\boldsymbol{A})} \\
(\boldsymbol{A}\boldsymbol{B})^{-1} &= \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}.
\end{aligned}$$

The Moore-Penrose inverse (pseudo-inverse) is a generalization of the inverse. It is the solution to the general least squares problem $\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$. For full-rank matrices, the left pseudo-inverse can be computed by

$$\boldsymbol{A}^+ = (\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top.$$

The left pseudo-inverse can be computed if $r = n$ and has the property $\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{I}_{n\times n}$.

The right pseudo-inverse can be computed as

$$\boldsymbol{A}^+ = A^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}.$$

The right pseudo-inverse can be computed if $r = m$ and has the property $\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{I}_{m\times m}$.

3.2 *Trace*

Definition 4 (Trace)**.** The trace of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is the sum of its diagonal,

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}.$$

Properties. Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{n\times n}, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, c, d \in \mathbb{R}$, then

$$\begin{aligned}
\operatorname{tr}(c\boldsymbol{A} + d\boldsymbol{B}) &= c \cdot \operatorname{tr}(\boldsymbol{A}) + d \cdot \operatorname{tr}(\boldsymbol{B}) \\
\operatorname{tr}(\boldsymbol{A}) &= \operatorname{tr}\left(\boldsymbol{A}^\top\right) \\
\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) &= \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) \\
\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) &= \operatorname{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{B})\operatorname{tr}(\boldsymbol{B}\boldsymbol{C}\boldsymbol{A}) \\
\operatorname{tr}\left(\boldsymbol{A}^\top\boldsymbol{B}\right) &= \operatorname{tr}\left(\boldsymbol{B}^\top\boldsymbol{A}\right) = \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{B}^\top\right) = \operatorname{tr}\left(\boldsymbol{B}\boldsymbol{A}^\top\right) \\
\boldsymbol{x}^\top\boldsymbol{y} &= \operatorname{tr}\left(\boldsymbol{x}^\top\boldsymbol{y}\right) = \operatorname{tr}\left(\boldsymbol{x}\boldsymbol{y}^\top\right).
\end{aligned}$$

Linearity.

Cyclic property.

$\boldsymbol{x}\boldsymbol{y}^\top$ is a rank-1 matrix.

Furthermore, the trace of a matrix is equal to the sum of the eigenvalues of the matrix.

3.3 *Orthogonal projection*

The projection $\boldsymbol{a}_1$ of a vector $\boldsymbol{a}$ on another vector $\boldsymbol{b}$ can be computed as

$$\boldsymbol{a}_1 = \frac{\boldsymbol{b}^\top \boldsymbol{a}}{\boldsymbol{b}^\top \boldsymbol{b}} \boldsymbol{b}.$$

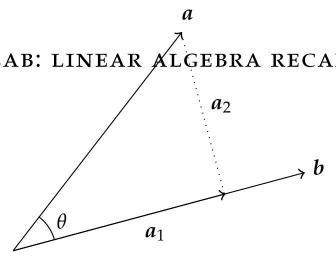

Figure 2. Projection of $\boldsymbol{a}$ on $\boldsymbol{b}$, denoted $\boldsymbol{a}_1$, and the rejection of $\boldsymbol{a}$ from $\boldsymbol{b}$, denoted $\boldsymbol{a}_2$.

The rejection can then be computed as

$$\boldsymbol{a}_2 = \boldsymbol{a} - \boldsymbol{a}_1.$$

We can also project a vector $\boldsymbol{a}$ onto the column space of a matrix $\boldsymbol{B} \in \mathbb{R}^{m \times n}$, denoted $\boldsymbol{a}_{\boldsymbol{B}}$. This can be computed by a matrix multiplication,

$$\boldsymbol{P} = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top,$$

which is equivalent to the left Moore-Penrose inverse.

Definition 5 (Projection matrix)**.** A square matrix $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ is called a projection matrix if it is idempotent, *i.e.*, $\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}$.

Properties. Let $\boldsymbol{P}$ be an orthogonal projection matrix, then $\boldsymbol{P} = \boldsymbol{P}^\top$, *i.e.*, $\boldsymbol{P}$ is symmetric. Furthermore, the eigenvalues of a projection matrix are all ones and zeros, because of idempotency.

3.4 *Special matrices*

Definition 6 (Orthogonal matrix)**.** An orthogonal matrix is an invertible matrix whose columns $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are all orthogonal to each other and of unit length, *i.e.*,

$$\begin{aligned} \boldsymbol{q}_i^\top \boldsymbol{q}_j &= 0 \quad \forall i \neq j \in [n] \\ \boldsymbol{q}_i^\top \boldsymbol{q}_i &= 1 \quad \forall i \in [n]. \end{aligned}$$

Properties. Let $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ be an orthogonal matrix,

$$\begin{aligned} \boldsymbol{Q}^\top &= \boldsymbol{Q}^{-1} \\ \langle \boldsymbol{x}, \boldsymbol{y} \rangle &= \langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y} \rangle. \end{aligned}$$