

Least Squares, Null space, Row space, Projectors

Computational Intelligence, Lecture 2

by Sergei Savin

Spring 2022

- Motivating questions
- Four Fundamental Subspaces
- Null space
 - ▶ Definition
 - ▶ Calculation
- Null space projection
- Closest element from a linear subspace
- Orthogonality, definition
- Projection
- Vectors in Null space, Row space
- Row and Null spaces in linear equations

LEAST SQUARES AT A GLANCE (1)

Consider the following problem: find the smallest 2-norm \mathbf{x} for which the equality $\mathbf{Ax} = \mathbf{y}$ has least residual. This is the *least squares problem*.

- Residual is defined as $\mathbf{e} = \mathbf{Ax} - \mathbf{y}$
- The idea that we need to find the smallest 2-norm \mathbf{x} and, at the same time, least residual \mathbf{e} seems to call for compromise. However, we will find that there exist a linear subspace of solutions with the exact same residual, and on that subspace we can find \mathbf{x} with the smallest norm.

LEAST SQUARES AT A GLANCE (2)

Minimum of $\|\mathbf{e}\|_2 = \|\mathbf{Ax} - \mathbf{y}\|_2$ is achieved at the same \mathbf{x} as minimum of $(\mathbf{Ax} - \mathbf{y})^\top (\mathbf{Ax} - \mathbf{y})$. With that in mind, let us find its extremum:

$$2\mathbf{A}^\top (\mathbf{Ax} - \mathbf{y}) = 0 \quad (1)$$

$$\mathbf{A}^\top \mathbf{Ax} = \mathbf{A}^\top \mathbf{y} \quad (2)$$

$$\mathbf{x} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y} \quad (3)$$

Thus we can define a *pseudoinverse*:

$$\mathbf{A}^+ = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \quad (4)$$

LEAST SQUARES AT A GLANCE (3)

Thus the least residual solution to $\mathbf{Ax} = \mathbf{y}$ is written as:

$$\mathbf{x} = \mathbf{A}^+ \mathbf{y} \quad (5)$$

We proved that this is the least-residual solution, we will prove that the solution is smallest norm (out of all solutions with the same residual) in the following lectures.

LEAST SQUARES AND CLOSEST ELEMENT

You are given an equation $\mathbf{Ax} = \mathbf{y}$. You want to find an \mathbf{x} that gets the value of \mathbf{y} as close as possible to \mathbf{y}^* in terms of the 2-norm.

We know that $\mathbf{x} = \mathbf{A}^+\mathbf{y}^*$ gives us the least residual solution. Multiplying it by \mathbf{A} we get:

$$\mathbf{y} = \mathbf{AA}^+\mathbf{y}^* \tag{6}$$

This is the value of \mathbf{y} that we can achieve, which is closest to \mathbf{y}^* .

FOUR FUNDAMENTAL SUBSPACES

One of the key ideas in Linear Algebra is that every linear operator has four fundamental subspaces:

- Null space
- Row space
- Column space
- Left null space

Our goal is to understand them. The usefulness of this concept is enormous.

NULL SPACE

Definition

Consider the following task: find all solutions to the system of equations $\mathbf{A}\mathbf{x} = \mathbf{0}$.

It can be re-formulated as follows: find all elements of the *null space* of \mathbf{A} .

Definition 1

Null space of \mathbf{A} is the set of all vectors \mathbf{x} that \mathbf{A} maps to $\mathbf{0}$

We will denote null space as $\mathcal{N}(\mathbf{A})$. In the literature, it is often denoted as $\ker(\mathbf{A})$ or $\text{null}(\mathbf{A})$.

NULL SPACE

Calculation

Now we can find all solutions to the system of equations $\mathbf{Ax} = \mathbf{0}$ by using functions that generate an orthonormal *basis* in the null space of \mathbf{A} . In MATLAB it is function `null`, in Python/Scipy - `null_space`:

- `N = null(A).`

- `N = scipy.linalg.null_space(A).`

That is it! Space of solutions of $\mathbf{Ax} = \mathbf{0}$ is the span of the columns of \mathbf{N} , and all solutions \mathbf{x}^* can be represented as $\mathbf{x}^* = \mathbf{Nz}$; for any \mathbf{z} we get a unique solution, and for any solution - a unique \mathbf{z} .

NULL SPACE PROJECTION

Local coordinates

Let \mathbf{N} be the orthonormal basis in the null space of matrix \mathbf{A} . Then, if a vector \mathbf{x} lies in the null space of \mathbf{A} , it can be represented as:

$$\mathbf{x} = \mathbf{N}\mathbf{z} \tag{7}$$

where \mathbf{z} are coordinates of \mathbf{x} in the basis \mathbf{N} .

However, there are vectors which not only are not lying in the null space of \mathbf{A} , but the closest vector to them in the null space is the zero vector.

CLOSEST ELEMENT FROM A LINEAR SUBSPACE

Let $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Its null space has orthonormal basis $\mathbf{N} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

■ $\begin{bmatrix} -2 \\ 0 \end{bmatrix} = -2\mathbf{N}$, $\begin{bmatrix} 10 \\ 0 \end{bmatrix} = 10\mathbf{N}$, - both are in the null space.

■ for $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ the closest vector in the null space is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

■ for $\mathbf{y} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ the closest vector in the null space is $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

ORTHOGONALITY, DEFINITION (1)

Definition

Any two vectors, \mathbf{x} and \mathbf{y} , whose dot product is zero are said to be *orthogonal* to each other.

Definition

Vector \mathbf{x} , whose dot product with any $\mathbf{y} \in \mathcal{L}$ is orthogonal to the subspace \mathcal{L}

Definition (equivalent)

If for a vector \mathbf{x} , the closest vector to it from a linear subspace \mathcal{L} is zero vector, \mathbf{x} is called orthogonal to the subspace \mathcal{L} .

ORTHOGONALITY, DEFINITION (2)

Definition

The space of all vectors \mathbf{y} , orthogonal to a linear subspace \mathcal{L} is called *orthogonal complement* of \mathcal{L} and is denoted as \mathcal{L}^\perp .

Definition (equivalent)

The space of all vectors \mathbf{y} , such that $\text{dot}(\mathbf{y}, \mathbf{x}) = 0, \forall \mathbf{x} \in \mathcal{L}$ is called *orthogonal complement* of \mathcal{L} .

Therefore $\mathbf{x} \in \mathcal{L}$ and $\mathbf{y} \in \mathcal{L}^\perp$ implies $\text{dot}(\mathbf{y}, \mathbf{x}) = 0$.

PROJECTION

Part 1

Let \mathbf{L} be an orthonormal basis in a linear subspace \mathcal{L} . Take vector $\mathbf{a} = \mathbf{x} + \mathbf{y}$, where \mathbf{x} lies in the subspace \mathcal{L} , and \mathbf{y} lies in the subspace \mathcal{L}^\perp .

Definition

We call such vector \mathbf{x} a *projection* of \mathbf{a} onto subspace \mathcal{L} , and such vector \mathbf{y} a projection of \mathbf{a} onto subspace \mathcal{L}^\perp .

Projection is the closest element in the subspace to a given vector. Projection of \mathbf{a} onto \mathcal{L} can be found as:

$$\mathbf{x} = \mathbf{L}\mathbf{L}^+ \mathbf{a} \tag{8}$$

Since \mathbf{L} is orthonormal, this is the same as $\mathbf{x} = \mathbf{L}\mathbf{L}^\top \mathbf{a}$

PROJECTION

Part 2

Since $\mathbf{a} = \mathbf{x} + \mathbf{y}$, and $\mathbf{x} = \mathbf{L}\mathbf{L}^\top \mathbf{a}$, we can write:

$$\mathbf{a} = \mathbf{L}\mathbf{L}^\top \mathbf{a} + \mathbf{y} \quad (9)$$

from which it follows that the projection of \mathbf{a} onto \mathcal{L}^\perp can be found as:

$$\mathbf{y} = (\mathbf{I} - \mathbf{L}\mathbf{L}^\top) \mathbf{a} \quad (10)$$

where \mathbf{I} is an identity matrix. Since \mathbf{L} is orthonormal, this is the same as $\mathbf{y} = (\mathbf{I} - \mathbf{L}\mathbf{L}^\top) \mathbf{a}$

ROW SPACE

Definition

Definition

Let \mathcal{N} be null space of \mathbf{A} . Then orthogonal subspace \mathcal{N}^\perp is called *row space* of \mathbf{A} .

Definition

Row space of \mathbf{A} is the space of all smallest-norm solutions of $\mathbf{A}\mathbf{x} = \mathbf{y}$, for $\forall \mathbf{y}$.

We will denote row space as \mathcal{R} .

VECTORS IN NULL SPACE, ROW SPACE

Given vector \mathbf{x} , matrix \mathbf{A} and its nulls space basis \mathbf{N} , and we check if \mathbf{x} is in the null space of \mathbf{A} . The simplest way is to check if $\mathbf{A}\mathbf{x} = 0$. But sometimes we may want to avoid computing $\mathbf{A}\mathbf{x}$, for example if the number of elements of \mathbf{A} is much larger than the number of elements of \mathbf{N} .

If \mathbf{x} is in the null space of \mathbf{A} , it will have zero projection onto the row space of \mathbf{A} . This gives us the condition we can check:

$$(\mathbf{I} - \mathbf{N}\mathbf{N}^\top)\mathbf{x} = 0 \quad (11)$$

By the same logic, condition for being in the row space is as follows:

$$\mathbf{N}\mathbf{N}^\top \mathbf{x} = 0 \quad (12)$$

ROW AND NULL SPACES IN LINEAR EQUATIONS

Part 1

Consider another task: find all solutions to the system of equations $\mathbf{Ax} = \mathbf{y}$.

Assume we have two solutions to the system: \mathbf{x}_1 and \mathbf{x}_2 . We know that $\mathbf{Ax}_1 = \mathbf{Ax}_2 = \mathbf{y}$, hence $\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$. In other words, the difference between any two solutions lies in the null space of \mathbf{A} .

On the other hand, let \mathbf{x}^* be a solution, and $\mathbf{x}^N \in \mathcal{N}(\mathbf{A})$ be a vector in the null space of \mathbf{A} . Then $\mathbf{x}^* + \mathbf{x}^N$ is also a solution, since $\mathbf{A}(\mathbf{x}^* + \mathbf{x}^N) = \mathbf{Ax}^* + \mathbf{Ax}^N = \mathbf{Ax}^* = \mathbf{y}$.

Therefore, the solution space is given by a single partial solution $\mathbf{x}^p \notin \mathcal{N}(\mathbf{A})$ and the whole null space of \mathbf{A} .

ROW AND NULL SPACES IN LINEAR EQUATIONS

Part 2

There are infinitely many ways to choose \mathbf{x}^p , since if $\mathbf{x}^p \notin \mathcal{N}(\mathbf{A})$, then $(\mathbf{x}^p + \mathbf{x}^N) \notin \mathcal{N}(\mathbf{A})$, if $\mathbf{x}^N \in \mathcal{N}(\mathbf{A})$. However:

Statement 1

The smallest-norm \mathbf{x}^p will lie in the row space of \mathbf{A} .

We can prove it by observing that there can be only one $\mathbf{x}^p \in \mathcal{R}(\mathbf{A})$ and adding to it any vector $\mathbf{x}^N \in \mathcal{N}(\mathbf{A})$ can only increase its magnitude, as \mathbf{x}^p and \mathbf{x}^N are orthogonal.

ROW AND NULL SPACES IN LINEAR EQUATIONS

Part 3

If we have \mathbf{x}^* , which is a solution to $\mathbf{A}\mathbf{x} = \mathbf{y}$, we can find the particular solution $\mathbf{x}^p \in \mathcal{R}(\mathbf{A})$ as a projection:

$$\mathbf{x}^p = (\mathbf{I} - \mathbf{N}\mathbf{N}^\top)\mathbf{x}^* \quad (13)$$

where \mathbf{N} is the null space basis for \mathbf{A} . Alternatively, we can simply find it as:

$$\mathbf{x}^p = \mathbf{A}^+\mathbf{y} \quad (14)$$

All solutions to $\mathbf{A}\mathbf{x} = \mathbf{y}$ are then given as:

$$\mathbf{x}^* = \mathbf{A}^+\mathbf{y} + \mathbf{N}\mathbf{z}, \quad \forall \mathbf{z} \quad (15)$$

Lecture slides are available via Moodle.

You can help improve these slides at:

github.com/SergeiSa/Computational-Intelligence-Slides-Spring-2022



Check Moodle for additional links, videos, textbook suggestions.