Least Squares, Null space, Row space, Projectors

Computational Intelligence, Lecture 2

by Sergei Savin

Spring 2022

Content

Least squares at a glance

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

Solution to the least squares problem is given by a pseudoinverse:

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{y} \tag{1}$$

Notice that, surprisingly, this solves both minimizations at the same time: find smallest $\mathbf{x}$ among all least-residual solutions.

You are given equation $\mathbf{A}\mathbf{x} = \mathbf{y}$. Assume that you want to find such $\mathbf{x}$ that $\mathbf{y}$ achieves the value as close as possible to $\mathbf{y}^*$.

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

$$\mathbf{y} = \mathbf{A}\mathbf{A}^{+}\mathbf{y}^* \tag{2}$$

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

One of the key ideas in the 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 understating is enormous.

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 $\mathrm{null}(\mathbf{A})$.

Now we can find all solutions to the system of equations $\mathbf{A}\mathbf{x} = \mathbf{0}$ 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 $\mathbf{A}\mathbf{x} = \mathbf{0}$ is the span of the columns of $\mathbf{N}$, and all solutions $\mathbf{x}^*$ can be represented as $\mathbf{x}^* = \mathbf{N}\mathbf{z}$; for any $\mathbf{z}$ we get a unique solution, and for any solution - a unique $\mathbf{z}$.

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:

$$\mathbf{x} = \mathbf{N}\mathbf{z} \tag{3}$$

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

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

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

Definition

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 $\mathcal{L}$. We denote it as $\mathbf{x} \in \mathcal{L}^\perp$.

Definition

(equivalent) A vector $\mathbf{x}$, orthogonal to all elements of the subspace $\mathcal{L}$ is called *orthogonal* to the subspace $\mathcal{L}$.

Definition

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

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 $\mathbf{y}$ is orthogonal to $\mathcal{L}$.

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 of $\mathbf{a}$ onto $\mathcal{L}$ can be found as:

$$\mathbf{x} = \mathbf{L}\mathbf{L}^{+}\mathbf{a} \tag{4}$$

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

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

$$\mathbf{a} = \mathbf{L}\mathbf{L}^\top \mathbf{a} + \mathbf{y} \tag{5}$$

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

$$\mathbf{y} = (\mathbf{I} - \mathbf{L}\mathbf{L}^\top)\mathbf{a} \tag{6}$$

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

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 $\forall \mathbf{y}$, plus the zero vector, which is included in all linear subspaces.

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

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 $\mathbf{Ax} = 0$. But sometimes we may want to avoid computing $\mathbf{Ax}$, for example if the number of elements of $\mathbf{A}$ is much bigger than the number of elements of $\mathbf{N}$.

We notice that if $\mathbf{x}$ is in the null space of $\mathbf{A}$, it will have zero projection onto the row space of $\mathbf{A}$. So, the condition is as follows:

$$(\mathbf{I} - \mathbf{N}\mathbf{N}^\top)\mathbf{x} = 0 \tag{7}$$

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

$$\mathbf{N}\mathbf{N}^\top\mathbf{x} = 0 \tag{8}$$

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

Assume we have two solutions to the system: $\mathbf{x}_1$ and $\mathbf{x}_2$. We know that $\mathbf{A}\mathbf{x}_1 = \mathbf{A}\mathbf{x}_2 = \mathbf{y}$, 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 $\mathbf{A}(\mathbf{x}^* + \mathbf{x}^N) = \mathbf{A}\mathbf{x}^* + \mathbf{A}\mathbf{x}^N = \mathbf{A}\mathbf{x}^* = \mathbf{y}$.

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 $\mathbf{A}$.

There are infinitely many ways to chose $\mathbf{x}^p$, 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.

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:

$$\mathbf{x}^p = (\mathbf{I} - \mathbf{N}\mathbf{N}^\top)\mathbf{x}^* \tag{9}$$

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

$$\mathbf{x}^p = \mathbf{A}^+\mathbf{y} \tag{10}$$

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

$$\mathbf{x}^* = \mathbf{A}^+\mathbf{y} + \mathbf{N}\mathbf{z}, \ \forall \mathbf{z} \tag{11}$$

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.