Subspaces

Computational Intelligence, Lecture 2

by Sergei Savin

Spring 2023

Least squares at a glance (1)

Consider the following problem: find $\mathbf{x}$ that minimizes $||\mathbf{A}\mathbf{x} - \mathbf{y}||_2$. This is the *least squares problem.*

- The value $\mathbf{e} = \mathbf{A}\mathbf{x} - \mathbf{y}$ is called residual.
- $||\cdot||_2$ is a second norm, where $||\mathbf{e}||_2 = \sqrt{\mathbf{e}^\top \mathbf{e}}$
- Least squares problem is about finding *least residual solution.*

Least squares at a glance (2)

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

$$\frac{d}{dt}\left((\mathbf{A}\mathbf{x} - \mathbf{y})^\top(\mathbf{A}\mathbf{x} - \mathbf{y})\right) = 0 \tag{1}$$

$$2\mathbf{A}^\top(\mathbf{A}\mathbf{x} - \mathbf{y}) = 0 \tag{2}$$

$$\mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{A}^\top\mathbf{y} \tag{3}$$

$$\mathbf{x} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{y} \tag{4}$$

Thus we can define a *pseudoinverse*:

$$\mathbf{A}^+ = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top \tag{5}$$

Least squares at a glance (3)

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

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{y} \tag{6}$$

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

Least squares at a glance (4)

Orthonormal case

Matrix $\mathbf{M}$ is orthonormal, meaning $\mathbf{M}^\top \mathbf{M} = \mathbf{I}$. Its pseudoinverse can be simplified:

$$\mathbf{M}^+ = (\mathbf{M}^\top \mathbf{M})^{-1} \mathbf{M}^\top = \mathbf{M}^\top \tag{7}$$

Then least squares solution to the equation $\mathbf{M}\mathbf{x} = \mathbf{y}$ can be found as:

$$\mathbf{x}_{LS} = \mathbf{M}^\top \mathbf{y} \tag{8}$$

Computing the residual

Given an equation $\mathbf{Ax} = \mathbf{y}$ and least squares solution $\mathbf{x}_{LS} = \mathbf{A}^{+}\mathbf{y}$, let us compute the residual $\mathbf{e} = \mathbf{Ax}_{LS} - \mathbf{y}$. We substitute the solution:

$$\mathbf{e} = \mathbf{AA}^{+}\mathbf{y} - \mathbf{y} \tag{9}$$

We find that:

- The residual can be found as $\mathbf{e} = (\mathbf{AA}^{+} - \mathbf{I})\mathbf{y}$.
- The closest $\mathbf{Ax}$ can get to $\mathbf{y}$ is $\mathbf{y}^{*} = \mathbf{AA}^{+}\mathbf{y}$.
- Later we will find that $\mathbf{AA}^{+}\mathbf{y}$ is a *projection* of $\mathbf{y}$ onto a *column space* of $\mathbf{A}$.

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.

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

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).`

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:

$$\mathbf{x} = \mathbf{N}\mathbf{z} \tag{10}$$

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

$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. 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 $\mathbf{y}$, whose dot product with any $\mathbf{x} \in \mathcal{L}$ is orthogonal to the subspace $\mathcal{L}$

Definition (equivalent, see Appendix A)

If for a vector $\mathbf{y}$, the closest vector to it from a linear subspace $\mathcal{L}$ is zero vector, $\mathbf{y}$ 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 $\mathrm{dot}(\mathbf{y}, \mathbf{x}) = 0$, $\forall \mathbf{x} \in \mathcal{L}$ is called *orthogonal complement* of $\mathcal{L}$.

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

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:

$$\mathbf{x} = \mathbf{L}\mathbf{L}^{+}\mathbf{a} \tag{11}$$

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

$$\mathbf{a} = \mathbf{L}\mathbf{L}^{+}\mathbf{a} + \mathbf{y} \tag{12}$$

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

$$\mathbf{y} = (\mathbf{I} - \mathbf{L}\mathbf{L}^{+})\mathbf{a} \tag{13}$$

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 complement $\mathcal{N}^{\perp}$ is called *row space* of $\mathbf{A}$.

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 $\mathcal{R}$.

Given vector $\mathbf{x}$, matrix $\mathbf{A}$ and its null space basis $\mathbf{N}$, 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:

$$(\mathbf{I} - \mathbf{N}\mathbf{N}^\top)\mathbf{x} = 0 \tag{14}$$

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

$$\mathbf{N}\mathbf{N}^\top\mathbf{x} = 0 \tag{15}$$

Consider the following task: find all linear combinations of the columns of $\mathbf{A}$: $\{\mathbf{y} : \ \mathbf{y} = \mathbf{A}\mathbf{x}\}$.

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

Definition - column space

Column space of $\mathbf{A}$ is the set of all outputs of the matrix $\mathbf{A}$, for all possible inputs

We will denote column space as $\mathcal{C}(\mathbf{A})$. In the literature, it is often called an *image* of $\mathbf{A}$.

Column space basis

The problem of finding orthonormal basis in the column space of a matrix is often called *orthonormalization* of that matrix. Hence in MATLAB and Python/Scipy the function that does it is called `orth`:

- `C = orth(A)`.
- `C = scipy.linalg.orth(A)`.

Let $\mathbf{A}$ be a square matrix, a map from $\mathbb{X} = \mathbb{R}^n$ to $\mathbb{Y} = \mathbb{R}^n$. Notice that if it has a non-trivial null space, it follows that multiple unique inputs are being mapped by it to the same output:

$$\begin{aligned} \mathbf{y} = \mathbf{A}\mathbf{x}_r = \mathbf{A}(\mathbf{x}_r + \mathbf{x}_n), \\ \mathbf{x}_r \in \mathcal{R}(\mathbf{A}) \\ \forall \mathbf{x}_n \in \mathcal{N}(\mathbf{A}) \end{aligned} \tag{16}$$

In fact, if null space of $\mathbf{A}$ has k dimensions, it implies that an k-dimensional subspace of $\mathbb{X}$ is mapped to a single element of $\mathbb{Y}$.

It follows that in this case the dimensionality of the column space could not exceed $n - k$.

Given vector $\mathbf{y}$ and matrix $\mathbf{A}$, let us find $\mathbf{y}_c$ - projection of $\mathbf{y}$ onto the column space of $\mathbf{A}$.

Since $\mathbf{y}_c \in \mathcal{C}(\mathbf{A})$, we can find such $\mathbf{x}$ that $\mathbf{A}\mathbf{x} = \mathbf{y}_c$; So, the problems is to minimize the residual $||\mathbf{y}_c - \mathbf{y}|| = ||\mathbf{A}\mathbf{x} - \mathbf{y}||$, which is least squares problem: $\mathbf{x} = \mathbf{A}^{+}\mathbf{y}$. So:

$$\mathbf{y}_c = \mathbf{A}\mathbf{A}^{+}\mathbf{y} \in \mathcal{C}(\mathbf{A}) \tag{17}$$

Remember that computing the pseudoinverse is based on SVD decomposition, same as finding a basis in the null space or the column space, so in terms of computational expense, all projections we discussed are similar.

Similarly we can define a projector onto the row space. Given vector $\mathbf{x}$ and matrix $\mathbf{A}$, let us find projector of $\mathbf{x}$ onto the row space of $\mathbf{A}$:

$$\mathbf{x}_r = \mathbf{A}^{+}\mathbf{A}\mathbf{x} \in \mathcal{R}(\mathbf{A}) \tag{18}$$

You can think of this in the following terms: first we find output $\mathbf{A}\mathbf{x}$, then we find the smallest norm vector that produces this same output, and this vector has to 1) have the same row space projection (because output is the same), 2) has to lie in the row space. Hence it is the row space projector of $\mathbf{x}$.

Notice that we implicitly used the fact that columns of $\mathbf{A}^{+}$ lie in the row space of $\mathbf{A}$. We will prove this fact later. Additionally, we will prove that row space of $\mathbf{A}$ is equivalent to the column space of $\mathbf{A}^{\top}$.

Left null space

The subspace, orthogonal to the column space is called *left null space.*

Definition

Space of all vectors $\mathbf{y}$ orthogonal to the columns of $\mathbf{A}$ is called *left null space*: $\mathbf{y}^\top \mathbf{A} = 0$

You can think of left null space as a space of vectors that not only cannot be produced by operator $\mathbf{A}$, but the closest vector to them that can be produced is the zero vector.

Notice that $\mathbf{y}^\top \mathbf{A} = 0$ implies $\mathbf{A}^\top \mathbf{y} = 0$, meaning that left null space of $\mathbf{A}$ is equivalent to the null space of $\mathbf{A}^\top$.

If we want to project vector $\mathbf{y}$ onto the left null space of $\mathbf{A}$, we project it onto the column space, and subtract the result from $\mathbf{y}$:

$$\mathbf{y}_l = (\mathbf{I} - \mathbf{A}\mathbf{A}^{+})\mathbf{y} \in \mathcal{C}^{\perp}(\mathbf{A}) \tag{19}$$

If $\mathbf{C}$ is an orthonormal basis in the column space of $\mathbf{A}$, the projection can be found the following way:

$$\mathbf{y}_l = (\mathbf{I} - \mathbf{C}\mathbf{C}^{\top})\mathbf{y} \in \mathcal{C}^{\perp}(\mathbf{A}) \tag{20}$$

Singular Value Decomposition

Given $\mathbf{A} \in \mathbb{R}^{n,m}$ we can find its Singular Value Decomposition (SVD):

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{L} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R}^\top \\ \mathbf{N}^\top \end{bmatrix} \tag{21}$$

$$\mathbf{A} = \mathbf{C}\mathbf{\Sigma}\mathbf{R}^\top \tag{22}$$

where $\mathbf{C}$, $\mathbf{L}$, $\mathbf{R}$ and $\mathbf{N}$ are column, left null, row and null space bases (orthonormal), $\mathbf{\Sigma}$ is the diagonal matrix of singular values. The singular values are positive and are sorted in the decreasing order.

Rank of the matrix is computed as the size of $\mathbf{\Sigma}$. Note that numeric tolerance applies when deciding if the singular value is non-zero.

Pseudoinverse $\mathbf{A}^+$ is computed as:

$$\mathbf{A}^+ = \begin{bmatrix} \mathbf{R} & \mathbf{N} \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C}^\top \\ \mathbf{L}^\top \end{bmatrix} \tag{23}$$

$$\mathbf{A}^+ = \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^\top \tag{24}$$

Note that this proves that $\mathbf{A}^+$ lies in the row space of $\mathbf{A}$.

Let's find SVD decomposition of a $\mathbf{A}^\top$:

$$\mathbf{A}^\top = \begin{bmatrix} \mathbf{C}_t & \mathbf{L}_t \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R}_t^\top \\ \mathbf{N}_t^\top \end{bmatrix} \tag{25}$$

Let us transpose it (remembering that transpose of a diagonal matrix the original matrix $\mathbf{\Sigma}_t^\top = \mathbf{\Sigma}_t$):

$$\mathbf{A} = \begin{bmatrix} \mathbf{R}_t & \mathbf{N}_t \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C}_t^\top \\ \mathbf{L}_t^\top \end{bmatrix} \tag{26}$$

Thus we can see that the row space of the original matrix $\mathbf{A}$ is the column space of the transpose $\mathbf{A}^\top$. And the left null space of the original matrix $\mathbf{A}$ is the null space of the transpose $\mathbf{A}^\top$.

Projectors (1)

Let's prove that $\mathbf{A}\mathbf{A}^+$ is equivalent to $\mathbf{C}\mathbf{C}^\top$:

$$\mathbf{A}\mathbf{A}^+ = \mathbf{C}\boldsymbol{\Sigma}\mathbf{R}^\top\mathbf{R}\boldsymbol{\Sigma}^{-1}\mathbf{C}^\top \tag{27}$$

$$\mathbf{A}\mathbf{A}^+ = \mathbf{C}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{C}^\top \tag{28}$$

$$\mathbf{A}\mathbf{A}^+ = \mathbf{C}\mathbf{C}^\top \tag{29}$$

Projectors (2)

Let's prove that $\mathbf{A}^{+}\mathbf{A}$ is equivalent to $\mathbf{R}\mathbf{R}^{\top}$:

$$\mathbf{A}^{+}\mathbf{A} = \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^{\top}\mathbf{C}\mathbf{\Sigma}\mathbf{R}^{\top} \tag{30}$$

$$\mathbf{A}^{+}\mathbf{A} = \mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{R}^{\top} \tag{31}$$

$$\mathbf{A}^{+}\mathbf{A} = \mathbf{R}\mathbf{R}^{\top} \tag{32}$$

Projectors (3)

Let us denote $\mathbf{P} = \mathbf{A}\mathbf{A}^{+}$. Let's prove that $\mathbf{P}\mathbf{P} = \mathbf{P}$:

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{C}\mathbf{\Sigma}\mathbf{R}^{\top}\mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^{\top}\mathbf{C}\mathbf{\Sigma}\mathbf{R}^{\top}\mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{C}^{\top} \tag{33}$$

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{C}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{C}^{\top} \tag{34}$$

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{C}\mathbf{C}^{\top} = \mathbf{A}\mathbf{A}^{+} \tag{35}$$

The same is true for $\mathbf{P} = \mathbf{A}^{+}\mathbf{A}$: we can prove that $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}^{+}\mathbf{A}$.

Projectors (4)

Let's prove that $\mathbf{P}^+ = \mathbf{P}$.
First, we find basis in the linear space where $\mathbf{P}$ projects onto: $\mathbf{C} = \mathrm{orth}(\mathbf{P})$, therefore $\mathbf{P} = \mathbf{C}\mathbf{C}^\top$.

$$\mathbf{P}\mathbf{P}^+ = \mathbf{I}, \tag{36}$$

$$\mathbf{C}\mathbf{C}^\top\mathbf{P}^+ = \mathbf{I}, \tag{37}$$

$$\mathbf{C}^\top\mathbf{P}^+ = \mathbf{C}^\top, \tag{38}$$

$$\mathbf{P}^+ = \mathbf{C}\mathbf{C}^\top, \tag{39}$$

$$\mathbf{P}^+ = \mathbf{P}, \quad \square \tag{40}$$

- Minimum Norm Solutions, Math 484: Nonlinear Programming, Mikhail Lavrov
- Orthogonality, Math 484: Nonlinear Programming, Mikhail Lavrov
- Data Driven Science & Engineering. Machine Learning, Dynamical Systems, and Control, Steven L. Brunton, J. Nathan Kutz, chapter Singular Value Decomposition (SVD)

Lecture slides are available via Github, links are on Moodle

You can help improve these slides at:
github.com/SergeiSa/Computational-Intelligence-Slides-Spring-2023

We have two definitions of orthogonality of a vector and a subspace:

1. Vector $\mathbf{y}$, whose dot product with any $\mathbf{x} \in \mathcal{L}$ is orthogonal to the subspace $\mathcal{L}$
2. If for a vector $\mathbf{y}$, the closest vector to it from a linear subspace $\mathcal{L}$ is zero vector, $\mathbf{y}$ is called orthogonal to the subspace $\mathcal{L}$.

Let us prove their equivalence. First we show that 1) implies 2). Let $\mathbf{L}$ be orthonormal basis in $\mathcal{L}$. To find the closest element $\mathbf{y}^*$ of $\mathcal{L}$ to $\mathbf{y}$, we need to solve the least squares problem $\mathbf{L}\mathbf{z} = \mathbf{y}$, and multiply the solution by $\mathbf{L}$:

$$\mathbf{z}_{LS} = \mathbf{L}^\top \mathbf{y} = \mathbf{0} \tag{41}$$

$$\mathbf{y}^* = \mathbf{L}\mathbf{z}_{LS} = \mathbf{L}\mathbf{L}^\top \mathbf{y} = \mathbf{0} \tag{42}$$

Second, let us prove that 2) implies 1). Given that $\mathbf{y}^* = \mathbf{L}\mathbf{z}_{LS} = \mathbf{L}\mathbf{L}^\top\mathbf{y} = \mathbf{0}$ we need to prove that $\mathbf{L}^\top\mathbf{y} = \mathbf{0}$. We start by multiplying the last equation by $\mathbf{L}^\top$:

$$\mathbf{L}\mathbf{L}^\top\mathbf{y} = \mathbf{0} \tag{43}$$

$$\mathbf{L}^\top\mathbf{L}\mathbf{L}^\top\mathbf{y} = \mathbf{0} \tag{44}$$

$$\mathbf{L}^\top\mathbf{y} = \mathbf{0} \qquad \text{since } \mathbf{L}^\top\mathbf{L} = \mathbf{I} \tag{45}$$