Least Squares and the Pseudoinverse

Kenton Lam

August 29, 2019

The goal of this is to explain how the (Moore-Penrose) pseudoinverse relates to the least squares problem. Specifically, the psuedoinverse allows one to compute the least-norm solution to the problem in all cases, and if a unique solution exists it will give you the unique solution.

Notation

$\mathbf{A} \in \mathbb{R}^{m \times n}$ denotes an $m \times n$ matrix and $\mathbf{A}^\dagger$ is its pseudoinverse. Range $\mathbf{A}$ is the vector space spanned by the columns of $\mathbf{A}$ (the column space). As a consequence, Range $\mathbf{A}^\top$ is the span of the rows of $\mathbf{A}$ (the row space). Null $\mathbf{A}$ is the null space of $\mathbf{A}$.

Fundamental Theorem of Linear Algebra

Theorem. *If $\mathbf{A} \in \mathbb{C}^{m \times n}$, then*

$$\operatorname{Null}(\mathbf{A}) = \operatorname{Range}(\mathbf{A}^*)^\perp \text{ and } \operatorname{Null}(\mathbf{A}^*) = \operatorname{Range}(\mathbf{A})^\perp$$

where $\mathbf{B}^\perp$ denotes the orthogonal complement of $\mathbf{B}$ in its containing vector space.

Note that from here on, we will only consider real matrices so $\mathbf{A}^* = \mathbf{A}^\top$.

1 Least Squares

The ordinary least squares problem is the following.

Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a vector $\boldsymbol{b} \in \mathbb{R}^n$, solve

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2.$$

The vector $\boldsymbol{x}$ is called the solution to this least squares problem. Note that this solution may not be unique.

1.1 Pseudoinverse

First, let's discuss the pseudoinverse. It can be shown that there exists a unique $\mathbf{A}^\dagger \in \mathbb{R}^{n\times m}$ satisfying the Moore-Penrose conditions:

1. $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$
2. $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$
3. $(\mathbf{A}\mathbf{A}^\dagger)^\top = \mathbf{A}\mathbf{A}^\dagger$ (so $\mathbf{A}\mathbf{A}^\dagger$ is symmetric)
4. $(\mathbf{A}^\dagger\mathbf{A})^\top = \mathbf{A}^\dagger\mathbf{A}$ (so $\mathbf{A}^\dagger\mathbf{A}$ is symmetric)

Firstly, we show that $\mathbf{A}\mathbf{A}^\dagger$ is a projection matrix:

$$(\mathbf{A}\mathbf{A}^\dagger)(\mathbf{A}\mathbf{A}^\dagger) = (\mathbf{A}\mathbf{A}^\dagger\mathbf{A})\mathbf{A}^\dagger = \mathbf{A}\mathbf{A}^\dagger.$$

Because it is symmetric, it must also be an orthogonal projection onto its range. Recall that if $\mathbf{P}$ is an orthogonal projection, then $\mathbf{I}-\mathbf{P}$ is a projection onto a subspace orthogonal to Range $\mathbf{P}$. These facts will be used later on.

Now, we claim that Range $\mathbf{A}^\dagger$ = Range $\mathbf{A}^\top$. Because Range $\mathbf{A}^\top \perp$ Null $\mathbf{A}$, this is equivalent to saying Range $\mathbf{A}^\dagger \perp$ Null $\mathbf{A}$. Suppose $\mathbf{A}^\dagger\boldsymbol{x} \in$ Range $\mathbf{A}^\dagger$ and $\boldsymbol{y} \in$ Null $\mathbf{A}$. Then, to show they are orthogonal,

$$\begin{aligned}
(\mathbf{A}^\dagger\boldsymbol{x})^\top\boldsymbol{y} &= \boldsymbol{x}^\top(\mathbf{A}^\dagger)^\top\boldsymbol{y} \\
&= \boldsymbol{x}^\top(\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger)^\top\boldsymbol{y} && \text{(using property of pseudoinverse)} \\
&= \boldsymbol{x}^\top((\mathbf{A}^\dagger\mathbf{A})\mathbf{A}^\dagger)^\top\boldsymbol{y} \\
&= \boldsymbol{x}^\top(\mathbf{A}^\dagger)^\top(\mathbf{A}^\dagger\mathbf{A})^\top\boldsymbol{y} \\
&= \boldsymbol{x}^\top(\mathbf{A}^\dagger)^\top\mathbf{A}^\dagger\mathbf{A}\boldsymbol{y} && \text{(because } \mathbf{A}^\dagger\mathbf{A} \text{ is symmetric)} \\
&= \mathbf{0} && \text{(because } \boldsymbol{y} \in \text{Null}\,\mathbf{A})
\end{aligned}$$

An identical process shows that Null $\mathbf{A}^\dagger \perp$ Range $\mathbf{A}$ and hence Null $\mathbf{A}^\dagger$ = Null $\mathbf{A}^\top$.

1.2 Solving least squares

The goal is to solve for $\boldsymbol{x}$ which minimises

$$\frac{1}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2.$$

Looking only at the norm, we can decompose $\boldsymbol{b}$ into two orthogonal vectors—one in $\operatorname{Null}\mathbf{A}^\top$ and one in $\operatorname{Range}\mathbf{A}$. This is done using the orthogonal projection $\mathbf{A}\mathbf{A}^\dagger$ we found earlier. It should be clear that this has not changed the equation.

$$\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = \|\mathbf{A}\boldsymbol{x} - (\mathbf{A}\mathbf{A}^\dagger + (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger))\boldsymbol{b}\|^2$$

Now, we expand and regroup terms to get

$$\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = \| \underbrace{(\mathbf{A}\boldsymbol{x} - \mathbf{A}\mathbf{A}^\dagger\boldsymbol{b})}_{\in \operatorname{Range}\mathbf{A}} + \underbrace{(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\boldsymbol{b}}_{\in \operatorname{Null}\mathbf{A}^\top} \|^2.$$

It is obvious why the left part is in $\operatorname{Range}\mathbf{A}$. For the right part, recall that $\mathbf{A}\mathbf{A}^\dagger$ is an orthogonal projection onto $\operatorname{Range}\mathbf{A}$ so $\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger$ projects onto the orthogonal complement of this which is the $\operatorname{Null}\mathbf{A}^\top$.

Because $\operatorname{Null}\mathbf{A}^\top \perp \operatorname{Range}\mathbf{A}$, we can use Pythagoras' theorem to split the norm into

$$\|(\mathbf{A}\boldsymbol{x} - \mathbf{A}\mathbf{A}^\dagger\boldsymbol{b}) + (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\boldsymbol{b}\|^2 = \|\mathbf{A}\boldsymbol{x} - \mathbf{A}\mathbf{A}^\dagger\boldsymbol{b}\|^2 + \|(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\boldsymbol{b}\|^2$$

The rightmost component is a constant so we can't change it. In fact, if $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ has no exact solution, it is precisely because this part is non-zero. For our purpose of least squares, we can ignore it because

$$\arg\min \|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = \arg\min \|\mathbf{A}\boldsymbol{x} - \mathbf{A}\mathbf{A}^\dagger\boldsymbol{b}\|^2$$

The value of the minimum is 0. It's important to note that this can be zero even when the original least squares is not 0. This is because $\mathbf{A}\mathbf{A}^\dagger$ is an orthogonal projection onto $\operatorname{Range}\mathbf{A}$, so there always exists an $\boldsymbol{x}$ such that $\mathbf{A}\boldsymbol{x} = \mathbf{A}\mathbf{A}^\dagger\boldsymbol{b}$ and the norm is 0, namely $\boldsymbol{x} = \mathbf{A}^\dagger\boldsymbol{b}$.

Contrast this with the original $\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|$. If $\boldsymbol{b} \notin \operatorname{Range}\mathbf{A}$, this will never be 0. However, we can minimise it which is what we aim to do with least squares.

1.3 The minimum norm solution

Finally, we want show that $\boldsymbol{x} = \mathbf{A}^\dagger\boldsymbol{b}$ is the solution with the smallest norm. Because $\boldsymbol{x} \in \operatorname{Range}\mathbf{A}^\top$, $\boldsymbol{x}$ has no components in the direction of the null space. This is shown by

$$\boldsymbol{x} = \mathbf{A}^\dagger \boldsymbol{b} = (\mathbf{A}^\dagger \mathbf{A})\mathbf{A}^\dagger \boldsymbol{b} + \underbrace{(\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\mathbf{A}^\dagger}_{=\mathbf{0}} \boldsymbol{b}.$$

If we add some component from the null space, we will get other solutions (because $\mathbf{A}\boldsymbol{x}$ will not change). These extra solutions can be expressed as

$$\boldsymbol{x} = \mathbf{A}^\dagger \boldsymbol{b} + \underbrace{(\mathbf{I} - \mathbf{A}^\dagger \mathbf{A})\boldsymbol{y}}_{\in \operatorname{Null} \mathbf{A}}, \qquad \text{for any } y \in \mathbb{R}^m.$$

Clearly, these solutions will have a larger norm than the original $\mathbf{A}^\dagger \boldsymbol{b}$, so we conclude $\mathbf{A}^\dagger \boldsymbol{b}$ is the solution with the smallest norm.

Corollary: If $\operatorname{Null} \mathbf{A} = \{\mathbf{0}\}$, then $\mathbf{A}^\dagger \boldsymbol{b} = \mathbf{A}^{-1} \boldsymbol{b}$ is the unique solution.