Least Squares and the Pseudoinverse

Kenton Lam

August 28, 2019

The goal of this is to explain how the (Moore-Penrose) pseudoinverse relates to the last squares problem. Specifically, the psuedoinverse allows one to compute the least-norm solution to the problem.

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.

1 Least Squares

The ordinary least squares problems 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 Normal Equation

Let

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

It can be shown that f is convex. So to minimise f, it suffices to find an $\boldsymbol{x}$ such that $\nabla f(\boldsymbol{x}) = 0$. We can expand f into

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{2}\|\mathbf{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \\ &= \frac{1}{2}(\mathbf{A}\boldsymbol{x} - \boldsymbol{b})^\top(\mathbf{A}\boldsymbol{x} - \boldsymbol{b}) \\ &= \frac{1}{2}\boldsymbol{x}^\top\mathbf{A}^\top\mathbf{A}\boldsymbol{x} - \boldsymbol{b}^\top\mathbf{A}\boldsymbol{x} + \frac{1}{2}\boldsymbol{b}^\top\boldsymbol{b} \end{aligned}$$

To compute the gradient of this function, we use the following result: for all $\boldsymbol{b} \in \mathbb{R}^p$ and $\mathbf{B} \in \mathbb{R}^{p\times p}$,

$$\begin{aligned} f(\boldsymbol{x}) = \langle \boldsymbol{b}, \boldsymbol{x}\rangle &\implies \nabla f(\boldsymbol{x}) = \boldsymbol{b} \\ f(\boldsymbol{x}) = \langle \boldsymbol{x}, \mathbf{B}\boldsymbol{x}\rangle &\implies \nabla f(\boldsymbol{x}) = (\mathbf{B} + \mathbf{B}^\top)\boldsymbol{x} \end{aligned}$$

We can rewrite $f(\boldsymbol{x})$ in terms of inner products to see that

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{2}\boldsymbol{x}^\top\mathbf{A}^\top\mathbf{A}\boldsymbol{x} - \boldsymbol{b}^\top\mathbf{A}\boldsymbol{x} + \frac{1}{2}\boldsymbol{b}^\top\boldsymbol{b} \\ &= \frac{1}{2}\langle \boldsymbol{x}, \mathbf{A}^\top\mathbf{A}\boldsymbol{x}\rangle - \langle \mathbf{A}^\top\boldsymbol{b}, \boldsymbol{x}\rangle + \frac{1}{2}\boldsymbol{b}^\top\boldsymbol{b} \end{aligned}$$

$$\begin{aligned} \implies \nabla f(\boldsymbol{x}) &= \frac{1}{2}(\mathbf{A}^\top\mathbf{A} + \mathbf{A}^\top\mathbf{A})\boldsymbol{x} - \mathbf{A}^\top\boldsymbol{b} \\ &= \mathbf{A}^\top\mathbf{A}\boldsymbol{x} - \mathbf{A}^\top\boldsymbol{b} \end{aligned}$$

Setting $\nabla f(\boldsymbol{x}) = 0$ gives us the normal equation for least squares,

$$\mathbf{A}^\top\mathbf{A}\boldsymbol{x} = \mathbf{A}^\top\boldsymbol{b}.$$

Thus, any $\boldsymbol{x}$ satisfying the normal equation will solve the initial least squares problem.

1.2 Pseudoinverse

We assume 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.

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

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

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

1.3 Conclusion

With this in mind, we return to the normal equation. By the rank-nullity theorem, we know that $\operatorname{Range}\mathbf{A}^\top \oplus \operatorname{Null}\mathbf{A} = \mathbb{R}^n$. Then, any solution $\boldsymbol{x}$ can be written uniquely as $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$ where $\boldsymbol{u} \in \operatorname{Range}\mathbf{A}^\top$ and $\boldsymbol{v} \in \operatorname{Null}\mathbf{A}$.

Moreover, because we defined $\boldsymbol{u} \in \operatorname{Range} \mathbf{A}^\dagger$, there must exist a unique[1] $\boldsymbol{y} \in \operatorname{Range}(\mathbf{A}^{\dagger\top}) = \operatorname{Range} \mathbf{A}$ such that $\boldsymbol{u} = \mathbf{A}^\dagger \boldsymbol{y}$. Thus,

$$\begin{aligned}
\mathbf{A}^\top \mathbf{A} \boldsymbol{x} &= \mathbf{A}^\top \boldsymbol{b} \\
\mathbf{A}^\top \mathbf{A} \mathbf{A}^\dagger \boldsymbol{y} &= \mathbf{A}^\top \boldsymbol{b} \\
\mathbf{A}^\top (\mathbf{A} \mathbf{A}^\dagger)^\top \boldsymbol{y} &= \mathbf{A}^\top \boldsymbol{b} \\
(\mathbf{A} \mathbf{A}^\dagger \mathbf{A})^\top \boldsymbol{y} &= \mathbf{A}^\top \boldsymbol{b} \\
\mathbf{A}^\top \boldsymbol{y} &= \mathbf{A}^\top \boldsymbol{b}
\end{aligned}$$

We have $\boldsymbol{y} \in \operatorname{Range} \mathbf{A}$ and $\mathbf{A}^\top \boldsymbol{b} \in \operatorname{Range} \mathbf{A}^\top$ (obviously). The matrix $\mathbf{A}^\top$ is a linear map which is injective when its domain is restricted to $\operatorname{Range} \mathbf{A}$. Therefore, the only possible $\boldsymbol{y}$ which satisfies this equation is $\boldsymbol{b}$.

Also, we said that any solution would be of the form

$$\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}$$

where $\boldsymbol{u} \in \operatorname{Range} \mathbf{A}^\top$ and $\boldsymbol{v} \in \operatorname{Null} \mathbf{A}$. We have shown $\boldsymbol{u} = \mathbf{A}^\dagger \boldsymbol{b}$ so the general solution to the least squares problem is given by

$$\boldsymbol{x} = \mathbf{A}^\dagger \boldsymbol{b} + \boldsymbol{v}, \quad \text{where } \boldsymbol{v} \in \operatorname{Null} \mathbf{A}.$$

An immediate consequence of this is that the solution is unique if and only if $\operatorname{Null} \mathbf{A} = \{\mathbf{0}\}$.

[1] when viewed as a linear map, a matrix is injective when restricted to its row space.