The Normal Equation

Dale Kim
Insomniac Games, Inc.
dkim@insomniacgames.com
@Roflraging

March 31, 2017

$$\boldsymbol{A^T A x} = \boldsymbol{A^T b}$$

1 Projection

Consider the problem of projecting a vector $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$, where the goal is to find some vector $\boldsymbol{p}$ in the column space of $\boldsymbol{A}$ that is as close as possible to $\boldsymbol{b}$. It is possible that $\boldsymbol{p}$ is different from $\boldsymbol{b}$, in which case we will use the difference as a measure of error:

$$\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$$

But what is $\boldsymbol{p}$? To answer this question, introduce a new vector $\boldsymbol{\hat{x}}$:

$$\boldsymbol{p} = \boldsymbol{A\hat{x}}$$

Because $\boldsymbol{p}$ must be in the column space of $\boldsymbol{A}$, we know there must exist some vector $\boldsymbol{\hat{x}}$ in the row space that produces it.

Directly from the problem description, the error needs to be minimized. Using some geometric reasoning, we can conclude that whatever $\boldsymbol{p}$ we are looking for should be orthogonal to $\boldsymbol{e}$, and that $\boldsymbol{e}$ must be in the left nullspace of $\boldsymbol{A}$:

$$\begin{aligned}
\boldsymbol{A^T e} &= 0 \\
\boldsymbol{A^T}(\boldsymbol{b} - \boldsymbol{p}) &= 0 \\
\boldsymbol{A^T}(\boldsymbol{b} - \boldsymbol{A\hat{x}}) &= 0 \\
\boldsymbol{A^T b} - \boldsymbol{A^T A\hat{x}} &= 0 \\
\boldsymbol{A^T A\hat{x}} &= \boldsymbol{A^T b}
\end{aligned}$$

At this point, the normal equation has been found, but how do you find the projection? Solve for $\boldsymbol{\hat{x}}$:

$$\begin{aligned}
\boldsymbol{A^T A\hat{x}} &= \boldsymbol{A^T b} \\
\boldsymbol{\hat{x}} &= (\boldsymbol{A^T A})^{-1}\boldsymbol{A^T b}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{p} &= \boldsymbol{A\hat{x}} \\
\boldsymbol{p} &= \boldsymbol{A}(\boldsymbol{A^T A})^{-1}\boldsymbol{A^T b}
\end{aligned}$$

Careful examination of the above equations will yield a new matrix $\boldsymbol{P}$ which is known as the projection matrix:

$$\begin{aligned}\boldsymbol{p} &= \boldsymbol{A}(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{b}\\ \boldsymbol{p} &= \boldsymbol{P}\boldsymbol{b}\\ \boldsymbol{P} &= \boldsymbol{A}(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\boldsymbol{T}}\end{aligned}$$

When you have $\boldsymbol{P}$, there is no need to compute $\hat{\boldsymbol{x}}$!

2 Least squares

Now consider the situation where a solution to $\boldsymbol{Ax} = \boldsymbol{b}$ does not exist. This is equivalent to saying that $\boldsymbol{b}$ is not in the column space of $\boldsymbol{A}$. Perhaps an approximate solution can be obtained? Intuitively, a good approximate solution $\hat{\boldsymbol{x}}$ would get as close to $\boldsymbol{b}$ as possible... I feel like I'm repeating myself.

In fact, the approximate solution is precisely $\hat{\boldsymbol{x}}$ that was found previously while looking at projection! The problems of projection and least squares are just two views of the normal equation. When the normal equation is viewed from the perspective of the row space, you're seeing the least squares approximation. But when viewed from the column space, it is the projection.

References

[1] Gilbert Strang. *Introduction to Linear Algebra, 4th Edition.* Welleslely - Cambridge Press, Wellesley, Massachusetts, 2009.