The 2 by 2 Real Pseudo Inverse

John Mount

2019-01-07

This is a brief note on the math behind the direct PRESS statistic calculation[1] found in the RcppDynProg package[2].

The actual ‘C++‘ code[3] is a bit ugly and intimidating. That is because we are using a verbose scalar notation to represent matrix concepts. In matrix notation we are solving a linear system by inverting a two by two matrix.[4] We are in turn inverting the two by two matrix by exploiting the following well know rule of how to invert a two by two matrix.

If $ad - bc$ is not zero then:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Throughout a, b, c, and d all real scalars.

This can be re-written as the following general relation.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = (ad - bc) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The above can be directly checked just by applying the rules for multiplying matrices to the left two matrices. This has a particularly pleasant presentation if we recognize that $ad - bc$ is the determinant of the left matrix and use the traditional vertical bar determinant notation.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Now there is an issue of what to do when $ad - bc$ is zero. For our implementation we apply Tikhonov regularization[5] which (barring the exact numeric coincidence of minus two times the expected value of the independent variable equaling our regularization constant) is going to be non-singular. For our actual

[1] `http://www.win-vector.com/blog/2014/09/estimating-generalization-error-with-the-press-statistic/`

[2] `https://github.com/WinVector/RcppDynProg`

[3] `https://github.com/WinVector/RcppDynProg/blob/master/src/xlin_fits.cpp`

[4] Yes, there are the usual admonitions that one should not invert a matrix to solve a linear system, but for systems this small they do not apply.

[5] `https://en.wikipedia.org/wiki/Tikhonov_regularization`

application, we could simply switch degenerate situations to the out-of sample mean implementation[6].

But, for fun, let's play with the math a bit.

There is an additional lesser known algebraic relation for two by two matrices.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = (a^2 + b^2 + c^2 + d^2) \begin{pmatrix} a & b \\ c & d \end{pmatrix} + (ad - bc) \begin{pmatrix} -d & c \\ b & -a \end{pmatrix}$$

Or (using transpose, matrix squared Frobenius norm, and determinant notation):

Theorem 1 *For any real 2 by 2 matrix* $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *we have:*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^\top \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \left\| \begin{matrix} a & b \\ c & d \end{matrix} \right\|_2^2 \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{pmatrix} -d & c \\ b & -a \end{pmatrix}$$

. The superscript "top" denoting the transpose operation, the $||.||_2^2$ *denoting sum of squares norm, and the single* $|.|$ *denoting determinant.*
□

This means, if $ad - bc$ is zero then:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^\top \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \left\| \begin{matrix} a & b \\ c & d \end{matrix} \right\|_2^2 \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Once we [confirm the above relation[7] we can also confirm that if $a^2 + b^2 + c^2 + d^2$ is not zero, then the following matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^+ = \frac{1}{a^2 + b^2 + c^2 + d^2} \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

satisfies all of the conditions for being the Moore-Penrose inverse[8] (or pseudo-inverse) of our original matrix. The superscript-plus denoting the Moore-Penrose inverse operation.

Or in transpose notation:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^+ = \frac{1}{a^2 + b^2 + c^2 + d^2} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^\top$$

This is called the pseudo-inverse because it acts like an inverse, even for non-invertible matrices. For A^+ to be a More-Penrose inverse we must confirm it obeys the following relations:

[6] https://github.com/WinVector/RcppDynProg/blob/master/src/const_costs.cpp

[7] https://github.com/WinVector/RcppDynProg/blob/master/extras/PseudoInverse.ipynb

[8] https://en.wikipedia.org/wiki/MoorePenrose_inverse

$$
\begin{aligned}
AA^+A &= A \\
A^+AA^+ &= A^+ \\
(AA^+)^\top &= AA^+ \\
(A^+A)^\top &= A^+A
\end{aligned}
$$

Theorem 1 lets us check the first relation, the other follow quickly as our A^+ is a simple scalar multiple of the transpose.[9]

All of the above check relations would be true for a classic inverse. We can think of A^+ as almost canceling a single A to the left or the A to the right.

Theorem 2 *For any real 2 by 2 matrix* $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$

The Moore-Penrose inverse is:

- *When* $ad - bc$ *is not zero:*

$$\frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

- *When* $ad - bc$ *is zero and* $a^2 + b^2 + c^2 + d^2$ *is not zero:*

$$\frac{1}{a^2 + b^2 + c^2 + d^2}\begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

- *Otherwise:*

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

.

□

For general matrices the situation is much more complicated.

The wealth of symmetries and relations is really kind of neat.

[9]With the same scaling for both A^+ and $(A^\top)^+$.