Left and right inverses; pseudoinverse

Although pseudoinverses will not appear on the exam, this lecture will help us to prepare.

Two sided inverse

A *2-sided inverse* of a matrix A is a matrix A^{-1} for which $AA^{-1} = I = A^{-1}A$. This is what we've called the *inverse* of A. Here $r = n = m$; the matrix A has full rank.

Left inverse

Recall that A has full column rank if its columns are independent; i.e. if $r = n$. In this case the nullspace of A contains just the zero vector. The equation $A\mathbf{x} = \mathbf{b}$ either has exactly one solution $\mathbf{x}$ or is not solvable.

The matrix A^TA is an invertible n by n symmetric matrix, so $(A^TA)^{-1}A^TA = I$. We say $A^{-1}_{\text{left}} = (A^TA)^{-1}A^T$ is a *left inverse* of A. (There may be other left inverses as well, but this is our favorite.) The fact that A^TA is invertible when A has full column rank was central to our discussion of least squares.

Note that AA^{-1}_{left} is an m by m matrix which only equals the identity if $m = n$. A rectangular matrix can't have a two sided inverse because either that matrix or its transpose has a nonzero nullspace.

left nullspace

Right inverse

If A has full row rank, then $r = m$. The nullspace of A^T contains only the zero vector; the rows of A are independent. The equation $A\mathbf{x} = \mathbf{b}$ always has at least one solution; the nullspace of A has dimension $n - m$, so there will be $n - m$ free variables and (if $n > m$) infinitely many solutions!

Matrices with full row rank have right inverses A^{-1}_{right} with $AA^{-1}_{\text{right}} = I$. The nicest one of these is $A^T(AA^T)^{-1}$. Check: A times $A^T(AAT)^{-1}$ is I.

Pseudoinverse

An invertible matrix ($r = m = n$) has only the zero vector in its nullspace and left nullspace. A matrix with full column rank $r = n$ has only the zero vector in its nullspace. A matrix with full row rank $r = m$ has only the zero vector in its left nullspace. The remaining case to consider is a matrix A for which $r < n$ and $r < m$.

If A has full column rank and $A^{-1}_{\text{left}} = (A^TA)^{-1}A^T$, then

$$AA^{-1}_{\text{left}} = A(A^TA)^{-1}A^T = P$$

is the matrix which projects $\mathbb{R}^m$ onto the column space of A. This is as close as we can get to the product $AM = I$.

Similarly, if A has full row rank then $A_{\text{right}}^{-1}A = A^T(AA^T)^{-1}A$ is the matrix which projects $\mathbb{R}^n$ onto the row space of A.

It's nontrivial nullspaces that cause trouble when we try to invert matrices. If $A\mathbf{x} = \mathbf{0}$ for some nonzero $\mathbf{x}$, then there's no hope of finding a matrix A^{-1} that will reverse this process to give $A^{-1}\mathbf{0} = \mathbf{x}$.

The vector $A\mathbf{x}$ is always in the column space of A. In fact, the correspondence between vectors $\mathbf{x}$ in the (r dimensional) row space and vectors $A\mathbf{x}$ in the (r dimensional) column space is one-to-one. In other words, if $\mathbf{x} \neq \mathbf{y}$ are vectors in the row space of A then $A\mathbf{x} \neq A\mathbf{y}$ in the column space of A. (The proof of this would make a good exam question.)

Proof that if $\mathbf{x} \neq \mathbf{y}$ then $A\mathbf{x} \neq A\mathbf{y}$

Suppose the statement is false. Then we can find $\mathbf{x} \neq \mathbf{y}$ in the row space of A for which $A\mathbf{x} = A\mathbf{y}$. But then $A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$, so $\mathbf{x} - \mathbf{y}$ is in the nullspace of A. But the row space of A is closed under linear combinations (like subtraction), so $\mathbf{x} - \mathbf{y}$ is also in the row space. The only vector in both the nullspace and the row space is the zero vector, so $\mathbf{x} - \mathbf{y} = \mathbf{0}$. This contradicts our assumption that $\mathbf{x}$ and $\mathbf{y}$ are not equal to each other.

We conclude that the mapping $\mathbf{x} \mapsto A\mathbf{x}$ from row space to column space is invertible. The inverse of this operation is called the *pseudoinverse* and is very useful to statisticians in their work with linear regression – they might not be able to guarantee that their matrices have full column rank $r = n$.

Finding the pseudoinverse A^+

The *pseudoinverse* A^+ of A is the matrix for which $\mathbf{x} = A^+A\mathbf{x}$ for all $\mathbf{x}$ in the row space of A. The nullspace of A^+ is the nullspace of A^T.

We start from the singular value decomposition $A = U\Sigma V^T$. Recall that Σ is a m by n matrix whose entries are zero except for the singular values $\sigma_1, \sigma_2, ..., \sigma_r$ which appear on the diagonal of its first r rows. The matrices U and V are orthonormal and therefore easy to invert. We only need to find a pseudoinverse for Σ.

The closest we can get to an inverse for Σ is an n by m matrix Σ^+ whose first r rows have $1/\sigma_1, 1/\sigma_2, ..., 1/\sigma_r$ on the diagonal. If $r = n = m$ then $\Sigma^+ = \Sigma^{-1}$. Always, the product of Σ and Σ^+ is a square matrix whose first r diagonal entries are 1 and whose other entries are 0.

If $A = U\Sigma V^T$ then its pseudoinverse is $A^+ = V\Sigma^+\mathbf{U}^T$. (Recall that $Q^T = Q^{-1}$ for orthogonal matrices U, V or Q.)

We would get a similar result if we included non-zero entries in the lower right corner of Σ^+, but we prefer not to have extra non-zero entries.

Conclusion

Although pseudoinverses will not appear on the exam, many of the topics we covered while discussing them (the four subspaces, the SVD, orthogonal matrices) are likely to appear.