LINEAR ALGEBRA. LECTURE 13

Similar matrices

Two square matrices A and B are *similar* if $B = M^{-1}AM$ for some matrix M. This allows us to put matrices into families in which all the matrices in a family are similar to each other. Then each family can be represented by a diagonal (or nearly diagonal) matrix.

Distinct eigenvalues

If A has a full set of eigenvectors we can create its eigenvector matrix S and write $S^{-1}AS = \Lambda$. So A is similar to Λ (choosing M to be this S).

If $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ then $\Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ and so A is similar to $\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$. But A is also similar to:

$$\begin{aligned} \overset{M^{-1}}{\begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix}} \overset{A}{\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}} \overset{M}{\begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}} &= \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 9 \\ 1 & 6 \end{bmatrix} \\ &= \overset{B}{\begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix}}. \end{aligned}$$

In addition, B is similar to Λ. All these similar matrices have the same eigenvalues, 3 and 1; we can check this by computing the trace and determinant of A and B.

Similar matrices have the same eigenvalues!

In fact, the matrices similar to A are all the 2 by 2 matrices with eigenvalues 3 and 1. Some other members of this family are $\begin{bmatrix} 3 & 7 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 7 \\ 0 & 3 \end{bmatrix}$. To prove that similar matrices have the same eigenvalues, suppose $A\mathbf{x} = \lambda\mathbf{x}$. We modify this equation to include $B = M^{-1}AM$:

$$\begin{aligned} AMM^{-1}\mathbf{x} &= \lambda\mathbf{x} \\ M^{-1}AMM^{-1}\mathbf{x} &= \lambda M^{-1}\mathbf{x} \\ BM^{-1}\mathbf{x} &= \lambda M^{-1}\mathbf{x}. \end{aligned}$$

The matrix B has the same λ as an eigenvalue. $M^{-1}\mathbf{x}$ is the eigenvector.

If two matrices are similar, they have the same eigenvalues and the same number of independent eigenvectors (but probably not the same eigenvectors).

When we diagonalize A, we're finding a diagonal matrix Λ that is similar to A. If two matrices have the same n distinct eigenvalues, they'll be similar to the same diagonal matrix.

The QR Algorithm for Computing Eigenvalues

The algorithm is almost magically simple. It starts with A_0, factors it by Gram-Schmidt into Q_0R_0, and then *reverses the factors*: $A_1 = R_0Q_0$. This new matrix A_1 is *similar* to the original one because $Q_0^{-1}A_0Q_0 = Q_0^{-1}(Q_0R_0)Q_0 = A_1$. So the process continues with no change in the eigenvalues:

$$\textbf{All } A_k \textbf{ are similar} \qquad A_k = Q_kR_k \quad \text{and then} \quad A_{k+1} = R_kQ_k.$$

This equation describes the *unshifted QR algorithm*, and almost always A_k approaches a triangular form, Its diagonal entries approach its eigenvalues, which are also the eigenvalues of A_0. If there was already some processing to obtain a tridiagonal form, then A_0 is connected to the absolutely original A by $Q^{-1}AQ = A_0$.

As it stands, the QR algorithm is good but not very good. To make it special, it needs two refinements: We must allow shifts to $A_k - \alpha_k I$, and we must ensure that the QR factorization at each step is very quick.

1. The Shifted Algorithm. If the number α_k is close to an eigenvalue, the step in equation (5) should be shifted immediately by α_k (which changes Q_k and R_k):

$$A_k = \alpha_k I = Q_k R_k \qquad \text{and then} \qquad A_{k+1} = R_k Q_k + \alpha_k I.$$

This matrix A_{k+1} is similar to A_k (always the same eigenvalues):

$$Q_k^{-1} A_k Q_k = Q_k^{-1}(Q_k R_k + \alpha_k I) Q_k = A_{k+1}.$$

What happens in practice is that the (n,n) entry of A_k—the one in the lower right-hand corner—is the first to approach an eigenvalue. That entry is the simplest and most popular choice for the shift α_k. Normally this produces quadratic convergence, and in the symmetric case even cubic convergence, to the smallest eigenvalue. After three or four steps of the shifted algorithm, the matrix A_k looks like this:

$$A_k = \left[\begin{array}{ccc|c} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ \hline 0 & 0 & \varepsilon & \lambda_1' \end{array}\right], \qquad \text{with} \quad \varepsilon \ll 1.$$

We accept the computed λ_1' as a very close approximation to the true λ_1. To find the next eigenvalue, the QR algorithm continues with the smaller matrix (3 by 3, in the illustration) in the upper left-hand corner. Its subdiagonal elements will be somewhat reduced by the first QR steps, and another two steps are sufficient to find λ_2. This gives a systematic procedure for finding all the eigenvalues. In fact, ***the QR method is now completely described***. It only remains to catch up on the eigenvectors—that is a single inverse power step—and to use the zeros that Householder created.

Singular value decomposition

The *singular value decomposition* of a matrix is usually referred to as the *SVD*. This is the final and best factorization of a matrix:

$$A = U\Sigma V^T$$

where U is orthogonal, Σ is diagonal, and V is orthogonal.

In the decomoposition $A = U\Sigma V^T$, A can be *any* matrix. We know that if A is symmetric positive definite its eigenvectors are orthogonal and we can write $A = Q\Lambda Q^T$. This is a special case of a SVD, with $U = V = Q$. For more general A, the SVD requires two different matrices U and V.

We've also learned how to write $A = S\Lambda S^{-1}$, where S is the matrix of n distinct eigenvectors of A. However, S may not be orthogonal; the matrices U and V in the SVD will be.

How it works

We can think of A as a linear transformation taking a vector $\mathbf{v}_1$ in its row space to a vector $\mathbf{u}_1 = A\mathbf{v}_1$ in its column space. The SVD arises from finding an orthogonal basis for the row space that gets transformed into an orthogonal basis for the column space: $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$.

It's not hard to find an orthogonal basis for the row space – the Gram-Schmidt process gives us one right away. But in general, there's no reason to expect A to transform that basis to another orthogonal basis.

You may be wondering about the vectors in the nullspaces of A and A^T. These are no problem – zeros on the diagonal of Σ will take care of them.

Matrix language

The heart of the problem is to find an orthonormal basis $\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_r$ for the row space of A for which

$$\begin{aligned} A\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix} &= \begin{bmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \cdots & \sigma_r\mathbf{u}_r \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}, \end{aligned}$$

with $\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_r$ an orthonormal basis for the column space of A. Once we add in the nullspaces, this equation will become $AV = U\Sigma$. (We can complete the orthonormal bases $\mathbf{v}_1, ...\mathbf{v}_r$ and $\mathbf{u}_1, ...\mathbf{u}_r$ to orthonormal bases for the entire space any way we want. Since $\mathbf{v}_{r+1}, ...\mathbf{v}_n$ will be in the nullspace of A, the diagonal entries $\sigma_{r+1}, ...\sigma_n$ will be 0.)

The columns of U and V are bases for the row and column spaces, respectively. Usually $U \neq V$, but if A is positive definite we can use the *same* basis for its row and column space!

Calculation

Suppose A is the invertible matrix $\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$. We want to find vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in the row space $\mathbb{R}^2$, $\mathbf{u}_1$ and $\mathbf{u}_2$ in the column space $\mathbb{R}^2$, and positive numbers σ_1 and σ_2 so that the $\mathbf{v}_i$ are orthonormal, the $\mathbf{u}_i$ are orthonormal, and the σ_i are the scaling factors for which $A\mathbf{v}_i = \sigma_i u_i$.

This is a big step toward finding orthonormal matrices V and U and a diagonal matrix Σ for which:

$$AV = U\Sigma.$$

Since V is orthogonal, we can multiply both sides by $V^{-1} = V^T$ to get:

$$A = U\Sigma V^T.$$

Rather than solving for U, V and Σ simultaneously, we multiply both sides by $A^T = V\Sigma^T U^T$ to get:

$$\begin{aligned} A^T A &= V\Sigma U^{-1} U\Sigma V^T \\ &= V\Sigma^2 V^T \\ &= V \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} V^T. \end{aligned}$$

This is in the form $Q\Lambda Q^T$; we can now find V by diagonalizing the symmetric positive definite (or semidefinite) matrix $A^T A$. The columns of V are eigenvectors of $A^T A$ and the eigenvalues of $A^T A$ are the values σ_i^2. (We choose σ_i to be the positive square root of λ_i.)

To find U, we do the same thing with AA^T.

SVD example

We return to our matrix $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$. We start by computing

$$\begin{aligned} A^T A &= \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}. \end{aligned}$$

The eigenvectors of this matrix will give us the vectors $\mathbf{v}_i$, and the eigenvalues will gives us the numbers σ_i.

Two orthogonal eigenvectors of $A^T A$ are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. To get an orthonormal basis, let $\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$. These have eigenvalues $\sigma_1^2 = 32$ and $\sigma_2^2 = 18$. We now have:

$$\begin{array}{ccccc} A & & U & \Sigma & V^T \\ \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} & = & \begin{bmatrix} & \\ & \end{bmatrix} & \begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix} & \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}. \end{array}$$

We could solve this for U, but for practice we'll find U by finding orthonormal eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$ for $AA^T = U\Sigma^2 U^T$.

$$AA^T = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix}$$
$$= \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}.$$

Luckily, AA^T happens to be diagonal. It's tempting to let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, as Professor Strang did in the lecture, but because $A\mathbf{v}_2 = \begin{bmatrix} 0 \\ -3\sqrt{2} \end{bmatrix}$ we instead have $\mathbf{u}_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ and $U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Note that this also gives us a chance to double check our calculation of σ_1 and σ_2.

Thus, the SVD of A is:

$$\begin{array}{cccc} A & U & \Sigma & V^T \\ \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = & \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} & \begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix} & \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}. \end{array}$$

Example with a nullspace

Now let $A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$. This has a one dimensional nullspace and one dimensional row and column spaces.

The row space of A consists of the multiples of $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$. The column space of A is made up of multiples of $\begin{bmatrix} 4 \\ 8 \end{bmatrix}$. The nullspace and left nullspace are perpendicular to the row and column spaces, respectively.

Unit basis vectors of the row and column spaces are $\mathbf{v}_1 = \begin{bmatrix} .8 \\ .6 \end{bmatrix}$ and $\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$. To compute σ_1 we find the nonzero eigenvalue of A^TA.

$$A^TA = \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$$
$$= \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}.$$

Because this is a rank 1 matrix, one eigenvalue must be 0. The other must equal the trace, so $\sigma_1^2 = 125$. After finding unit vectors perpendicular to $\mathbf{u}_1$ and $\mathbf{v}_1$ (basis vectors for the left nullspace and nullspace, respectively) we see that the SVD of A is:

$$\begin{array}{cccc} \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = & \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} & \begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix}. \\ A & U & \Sigma & V^T \end{array}$$

The singular value decomposition combines topics in linear algebra ranging from positive definite matrices to the four fundamental subspaces.

$\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_r$	is an orthonormal basis for the row space.
$\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_r$	is an orthonormal basis for the column space.
$\mathbf{v}_{r+1}, \ldots \mathbf{v}_n$	is an orthonormal basis for the nullspace.
$\mathbf{u}_{r+1}, \ldots \mathbf{u}_m$	is an orthonormal basis for the left nullspace.

These are the "right" bases to use, because $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$.

Left and right inverses; pseudoinverse

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

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_{right}^{-1} with $AA_{\text{right}}^{-1} = I$. 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 $A_{\text{left}}^{-1} = (A^TA)^{-1}A^T$, then

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

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.