9
Matrix Decompositions

9.1 Eigendecomposition

Let's discuss a square, $n \times n$ matrix $\mathbf{A}$. Provided $\mathbf{A}$ is not defective, it has n linearly independent eigenvectors which we will call $\mathbf{v}_1, \ldots, \mathbf{v}_n$. The eigenvectors are linearly independent and therefore form a basis for $\mathbb{R}^n$ (an *eigenbasis*). We said in the last chapter that any vector $\mathbf{x}$ can be decomposed onto the eigenbasis by finding coefficients $a_1, \ldots, a_n$ such that

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$$

Multiplying the vector $\mathbf{x}$ by the matrix $\mathbf{A}$ is equivalent to scaling each term in the decomposition by the corresponding eigenvalue (λ_i).

$$\mathbf{A}\mathbf{x} = a_1\lambda_1\mathbf{v}_1 + a_2\lambda_2\mathbf{v}_2 + \cdots + a_n\lambda_n\mathbf{v}_n$$

We can think of matrix multiplication as a transformation with three steps.

1. Decompose the input vector onto the eigenbasis of the matrix.

2. Scale each term in the decomposition by the appropriate eigenvalue.

3. Reassembly, or "un-decompose" the output vector.

Each of these steps can be represented by a matrix operation. First, we collect the n eigenvalue into a matrix $\mathbf{V}$.

$$\mathbf{V} = (\mathbf{v}_1\,\mathbf{v}_2\,\cdots\,\mathbf{v}_n)$$

Each column in the matrix $\mathbf{V}$ is an eigenvector of the matrix $\mathbf{A}$. To decompose the vector $\mathbf{x}$ onto the columns of $\mathbf{V}$ we find the coefficients $a_1, \ldots, a_n$ by solving the linear system

$$\mathbf{V}\mathbf{a} = \mathbf{x}$$

where $\mathbf{a}$ is a vector holding the coefficients $a_1, \ldots, a_n$. The matrix $\mathbf{V}$ is square and has linearly independent columns (the eigenvectors of $\mathbf{A}$), so its inverse exists. The coefficients for decomposing the vector $\mathbf{x}$ onto the eigenbasis of the matrix $\mathbf{A}$ are

$$\mathbf{a} = \mathbf{V}^{-1}\mathbf{x}$$

If the inverse matrix $\mathbf{V}^{-1}$ decomposes a vector into a set of coefficients $\mathbf{a}$, then multiplying the coefficients vector $\mathbf{a}$ by the original matrix must reassemble the vector $\mathbf{x}$. Looking back at the three steps we defined above, we can use multiplication by $\mathbf{V}^{-1}$ to complete step 1 and multiply by $\mathbf{V}$ to perform step 3. For step 2, we need to scale the individual coefficients by the appropriate eigenvalues. We define a scaling matrix $\mathbf{\Lambda}$ as a diagonal matrix of the eigenvalues:

In other words, if $\mathbf{V}^{-1}$ decomposes a vector, $(\mathbf{V}^{-1})^{-1} = \mathbf{V}$ must undo the decomposition.

We use the uppercase Greek lambda ($\mathbf{\Lambda}$) to denote the matrix of eigenvalues λ_i (lowercase lambda).

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Notice that

$$\mathbf{\Lambda a} = \begin{pmatrix} \lambda_1 a_1 \\ \lambda_2 a_2 \\ \vdots \\ \lambda_n a_n \end{pmatrix}$$

so the matrix $\mathbf{\Lambda}$ scales the ith entry of the input vector by the ith eigenvalue.

We now have matrix operations for decomposing onto an eigenbasis ($\mathbf{V}^{-1}$), scaling by eigenvalues ($\mathbf{\Lambda}$), and reassembling the output vector ($\mathbf{V}$). Putting everything together, we see that matrix multiplication can be expressed as an *eigendecomposition* by

Eigendecomposition is the last time we will use the prefix "eigen-". Feel free to use it on other everyday words to appear smarter.

$$\mathbf{Ax} = \mathbf{V\Lambda V}^{-1}\mathbf{x}$$

Equivalently, we can say that the matrix $\mathbf{A}$ itself can be as the product of three matrices ($\mathbf{A} = \mathbf{V\Lambda V}^{-1}$) if $\mathbf{A}$ has a complete set of eigenvectors. There are two ways to interpret the dependence on a complete set of eigenvectors. Viewed technically, the matrix $\mathbf{V}$ can only be inverted if it is full rank, so $\mathbf{V}^{-1}$ does not exist if one or more eigenvectors is missing. More intuitively, the eigendecomposition defines a unique mapping between the input and output vectors. Uniqueness requires a basis, since a vector decomposition is only unique if the set of vectors form a basis. If the matrix $\mathbf{A}$ is defective, its eigenvectors do not form an eigenbasis and there cannot be a unique mapping between inputs and outputs.

9.2 *Singular Value Decomposition*

The eigendecomposition is limited to square matrices with a complete set of eigenvectors. However, the idea that matrices can be factored into three operations (decomposition, scaling, and reassembly) generalizes to all matrices, even non-square matrices. The generalized equivalent of the eigendecomposition is called the *Singular Value Decomposition*, or (SVD).

Singular Value Decomposition. *Any $m \times n$ matrix $\mathbf{A}$ can be factored into the product of three matrices*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$$

where

- $\mathbf{U}$ *is an orthogonal $m \times m$ matrix.*
- $\mathbf{\Sigma}$ *is a diagonal $m \times n$ matrix with nonzero entries.*
- $\mathbf{V}$ *is an orthogonal $n \times n$ matrix.*

The square matrices $\mathbf{U}$ and $\mathbf{V}$ are *orthogonal*, i.e. their columns form an orthonormal set of basis vectors. As we discussed previously, the inverse of an orthogonal matrix is simply its transpose, so $\mathbf{U}^{-1} = \mathbf{U}^{\mathrm{T}}$ and $\mathbf{V}^{-1} = \mathbf{V}^{\mathrm{T}}$. The $\mathbf{V}^{\mathrm{T}}$ term in the decomposition has the same role as the $\mathbf{V}^{-1}$ matrix in an eigendecomposition – projection of the input vector onto a new basis. The matrix $\mathbf{U}$ in SVD reassembles the output vector analogous to the vector $\mathbf{V}$ in an eigendecomposition.

If the entries in $\mathbf{A}$ were complex numbers, the matrices $\mathbf{U}$ and $\mathbf{V}$ would be *unitary*. The inverse of a unitary matrix is the complex conjugate of the matrix transpose.

The matrix $\mathbf{\Sigma}$ is diagonal but not necessarily square. It has the same dimensions as the original matrix $\mathbf{A}$. For a 3×5 matrix, the $\mathbf{\Sigma}$ has the form

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$$

If the matrix $\mathbf{A}$ was 5×3, we would have

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The elements along the diagonal of $\mathbf{\Sigma}$ are called *singular values*. If $\mathbf{A}$ is an $m \times n$ matrix, the maximum number of nonzero singular values is $\min\{m, n\}$. The are the analogues of eigenvalues for non-square matrices. However, the singular values for a square matrix are not equal to the eigenvalues of the same matrix. Singular values are always nonnegative. If we arrange $\mathbf{\Sigma}$ such that the singular values are in descending order, the SVD of a matrix is unique.

The columns in $\mathbf{U}$ and $\mathbf{V}$ are called the left and right *singular vectors*, respectively. Just as there is a relationship between eigenvalues and eigenvectors, the columns in $\mathbf{U}$ and $\mathbf{V}$ are connected by the singular values in $\mathbf{\Sigma}$. If $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$, then

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$$

and

$$\mathbf{A}\mathbf{u}_i = \sigma_i \mathbf{v}_i$$

where $\mathbf{v}_i$ is ith right singular vector (the ith column in $\mathbf{V}$); $\mathbf{u}_i$ is the ith left singular vector (the ith column in $\mathbf{U}$); and σ_i is the ith singular value (the ith nonzero on the diagonal of $\mathbf{\Sigma}$).

Applications of the SVD

Rank of a matrix

The rank of a matrix $\mathbf{A}$ is equal to the number of nonzero singular values (the number of nonzero values along the diagonal of $\mathbf{\Sigma}$). This is true for both square and nonsquare matrices. Notice that the way we defined the diagonal matrix $\mathbf{\Sigma}$ implies that the number of singular values must be at most $\min\{m, n\}$ for an $m \times n$ matrix. This requirement agrees with our knowledge that $\mathrm{rank}(\mathbf{A}) \leq \min\{m, n\}$.

The pseudoinverse of a matrix

Our definition of a matrix inverse applies only to square matrices. For nonsquare matrices we can use the SVD to construct a pseudoinverse. We represent the pseudoinverse of a matrix $\mathbf{A}$ as $\mathbf{A}^+$. We simply reverse and invert the factorization of $\mathbf{A}$, i.e.

$$\mathbf{A}^+ = \left(\mathbf{V}^{\mathrm{T}}\right)^{-1}\mathbf{\Sigma}^+\mathbf{U}^{-1}$$

We can simplify this expression with knowledge that $\mathbf{V}$ and $\mathbf{U}$ are orthogonal, so $\left(\mathbf{V}^{\mathrm{T}}\right)^{-1} = \left(\mathbf{V}^{\mathrm{T}}\right)^{\mathrm{T}} = \mathbf{V}$ and $\mathbf{U}^{-1} = \mathbf{U}^{\mathrm{T}}$. Thus

$$\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}$$

The matrix $\mathbf{\Sigma}^+$ is the pseudoinverse of the diagonal matrix $\mathbf{\Sigma}$. This is simply the transpose of $\mathbf{\Sigma}$where all entries of the diagonal are replaced by their multiplicative inverse. For a 3×5 matrix $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$$

the pseudoinverse $\mathbf{\Sigma}^+$ is

$$\mathbf{\Sigma}^+ = \begin{pmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & 1/\sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Some properties of the pseudoinverse:

- If $\mathbf{A}^+$ exists, then $\mathbf{A}^+\mathbf{A} = \mathbf{I}$.
- If a matrix $\mathbf{A}$ is full rank, then $\mathbf{A}^+ = \left(\mathbf{A}^\mathrm{T}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}$. This alternative formulation requires inverting a matrix, so the SVD form above is a far better method for finding a pseudoinverse.
- If $\mathbf{A}$ is square and invertible, then $\mathbf{A}^{-1} = \mathbf{A}^+$, i.e. the pseudoinverse is equal to the normal inverse.
- It is always true that $(\mathbf{A}^+)^+ = \mathbf{A}$, just as for square, invertible matrices.
- If a pseudoinverse exists, then it is unique for a given matrix.