INVERSE OF A MATRIX. SINGULAR VALUE DECOMPOSITION

1. Inverse of a Matrix

We have seen how we can solve $\mathbf{A}x = b$ using Elimination. In some cases, we can simply solve the system of equations by taking the inverse of the matrix $\mathbf{A}$. An inverse of a matrix satisfies the equation: $\mathbf{A^{-1}A} = \mathbf{I}$.

As you can guess, only square matrices can have inverses and not all square matrices have inverses. In order for a square matrix to have an inverse, its determinant should be non-zero. That is, $det(\mathbf{A}) \neq 0$. Further, once we know the inverse of a matrix, we can solve the system of equations as $x = \mathbf{A^{-1}}b$.

1.1. **Co-factors of a Matrix.** Before we can define inverses, we need to understand co-factors of a matrix as they play a special role in computing the inverses. If you recall, when we were computing the determinant of a matrix, we would start with the top row of the matrix and take one element at a time from the top-row and then compute the determinant of a sub-matrix which was left behind after the top row and the corresponding column was removed from the original matrix.

A co-factor is the determinant of this sub-matrix along with the appropriate sign (+ or -) that goes with it. For instance, if you have a matrix $\mathbf{A}$ as shown below:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{1}$$

If you remove the first row and first column, you get a sub-matrix, $\mathbf{M_{11}}$, which looks as below:

$$M = \begin{bmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{2}$$

In general, the sub-matrix $\mathbf{M_{ij}}$ has its ith row and jth column removed. When we computed the determinant of $\mathbf{A}$, we expressed it in terms of its sub-matrices as follows:

$$det(A) = a_{11}det(M_{11}) - a_{12}det(M_{12}) + a_{13}det(M_{13}) \cdots + (-1)^{1+n}a_{1n}det(M_{1n}) \tag{3}$$

We had an alternating pattern of + and - in front of each of the elements in the first row. In general, that pattern can be expressed as $(-1)^{i+j}$ when we are dealing with the ith row and jth column.

Formally a co-factor C_{ij} is defined as the determinant of the sub-matrix M_{ij} together with the appropriate sign $(-1)^{i+j}$, or $(-1)^{i+j}det(M_{ij})$. This simplifies the formula for the determinant of the matrix to:

$$det(A) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \cdots + a_{1n}C_{1n} \tag{4}$$

where C_{ij} are the co-factors. We only use the co-factors for the first row when computing the determinant. However, determinants can also be computed by starting at any row and taking the appropriate co-factors for that row. So, the general formula for the determinant of A reduces to

$$det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + a_{i3}C_{i3} \cdots + a_{in}C_{in} \tag{5}$$

1.2. **Inverse of a matrix.** Let's define a co-factor matrix whose elements are co-factors corresponding to a single row and column of the parent matrix $\mathbf{A}$. For instance if you have a 3x3 matrix $\mathbf{A}$ defined as

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{6}$$

Then the co-factor matrix corresponding to this is defined as

$$C = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \tag{7}$$

where each of the C_{ij} represent the corresponding co-factors of $\mathbf{A}$. Let us take the product of $\mathbf{A}$ and $\mathbf{C^T}$, the transpose of $\mathbf{C}$. We get,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \times \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \begin{bmatrix} det(A) & 0 & 0 \\ 0 & det(A) & 0 \\ 0 & 0 & det(A) \end{bmatrix} \tag{8}$$

We can easily verify that the diagonal elements of this product are $det(A)$. For instance, row 1 of $\mathbf{A}$ times column 1 of $\mathbf{C^T}$ is $a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = det(A)$ by the definition of the determinant. Let us now show that the off-diagonal elements are 0.

If you take row 2 of $\mathbf{A}$ time column 1 of $\mathbf{C^T}$, you get $a_{21}C_{11} + a_{22}C_{12} + a_{23}C_{13}$. This can be seen as the determinant of a new matrix $\mathbf{A}^*$ whose first two rows are a copy of the 2nd row of $\mathbf{A}$. You know that the first row of $\mathbf{A}^*$ is a copy of the 2nd row of $\mathbf{A}$ by construction and since $\mathbf{A}^*$ has the same first-row co-factors as $\mathbf{A}$, its remaining rows must be identical to that of $\mathbf{A}$. Therefore, it has two identical rows. Since $\mathbf{A}^*$ has two identical rows, it is not full-rank. It has to have a determinant that is equal to 0.

Now, we can see that $\mathbf{A^{-1}} = \mathbf{C^T}/det(\mathbf{A})$. So, we have a simple formula for constructing the inverse of a full-rank matrix. We compute its determinant and all its co-factors. We take the transpose of the co-factor matrix and divide it by the determinant of the matrix to get the inverse.

2. Singular Value Decomposition

SVD or Singular Value Decomposition is a highlight of Linear Algebra. As we have seen, many matrices are not invertible and in many instances, they are not even square. SVD can be employed in all these situations. It works for any $m \times n$ matrix of rank r. But, before we visit SVD, we need to spend some time with symmetric matrices and their eigenvalues. They will be needed for computing SVD.

2.1. **Symmetric matrices.** Symmetric matrices are matrices such that $A = A^T$ where the transpose of the matrix is the matrix itself. When A is symmetric, does $Ax = \lambda x$ have any special properties? Let's take another look at Matrix factorization, but this time with eigenvalues and eigenvectors instead of LDU. We'll see that with symmetric matrices, eigenvalues and eigenvectors take special forms which become apparent when we look at the factorized form of the matrix in terms of its eigenvalues and eigenvectors.

2.1.1. *Diagonalizing matrices using eigenvectors.* If the $n \times n$ square matrix $\mathbf{A}$ is full-rank (that is, has n linearly independent rows (or columns)) it will have n linearly independent eigenvectors. If we stack the eigenvectors into the columns, we'll get an $n \times n$ eigenvector matrix, $\mathbf{S}$.

Let's look at the product $\mathbf{A} \times \mathbf{S}$:

$$\mathbf{A} \times \mathbf{S} = \mathbf{A} \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix} = \begin{bmatrix} \lambda_1 e_1 & \lambda_2 e_2 & \cdots & \lambda_n e_n \end{bmatrix} \tag{9}$$

Since $\mathbf{A}$ times the first eigenvector e_1 is $\lambda_1 e_1$ and so forth. This can be viewed as the product of the $\mathbf{S}$ matrix and a diagonal matrix Λ whose diagonal entries are the eigenvalues. That is,

$$\mathbf{A} \times \mathbf{S} = \mathbf{S} \times \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \mathbf{S} \times \mathbf{\Lambda} \tag{10}$$

We can now represent $\mathbf{A}$ in terms of its factors comprised of eigenvalues and eigenvectors and vice-versa. $\mathbf{AS} = \mathbf{S\Lambda} \implies \mathbf{S^{-1}AS} = \mathbf{\Lambda}$ which we obtain by pre-multiplying both sides by $\mathbf{S^{-1}}$. Similarly, $\mathbf{A} = \mathbf{S\Lambda S^{-1}}$ by post-multiplying both sides by $\mathbf{S^{-1}}$.

Given a full-rank matrix $\mathbf{A}$, we can factorize it using eigenvectors and eigenvalues as shown above. Now, let's consider the case of a symmetric matrix bfA. We know that $\mathbf{A}$ can be factorized as $\mathbf{S\Lambda S^{-1}}$. Looking at $\mathbf{A^T}$, the transpose, we can show that $\mathbf{A^T} = (\mathbf{S\Lambda S^{-1}})^\mathbf{T} = (\mathbf{S^{-1}})^\mathbf{T}\mathbf{\Lambda S^T}$ by the rule for taking transposes of products of matrices. Since $\mathbf{A^T} = \mathbf{A}$, we can see that $\mathbf{S^{-1}} = \mathbf{S^T}$ by examination. This implies $\mathbf{S^TS} = \mathbf{I}$, or the eigenvectors of a symmetric matrix are orthogonal. Since we customarily choose the eigenvectors to be of unit length, for symmetric matrices, we end up with orthonormal eigenvectors, or a basis set in n-dimensional space.

Further, we can show that real symmetric matrices have real eigenvalues. This is known as the *Spectral Theorem.* For proof, please see Section 6.4 of the Strang textbook.

2.2. **Nullspace of a matrix.** We have seen how to solve a system of equations: $\mathbf{A}x = b$. Let's consider the special case where $b = 0$. What vectors x solve the equation $\mathbf{A}x = 0$?

The set of vectors x that satisfy the above equation represent the null space of $\mathbf{A}$. Notice that this is a proper subspace as all the vector space properties can be verified to hold in this space. If $\mathbf{A}$ is an $n \times n$ matrix, the solution vectors x have n components and the vectors x live in $\mathbf{R^n}$. The null space of matrix $\mathbf{A}$ is a proper subspace of $\mathbf{R^n}$.

Let us start with a non-invertible (singular) matrix and see what its null space looks like:

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \tag{11}$$

As you can see from the trivial example, the second row is just a multiple of the first row. Therefore, we get the following two equations that need to be solved:

$$\begin{aligned} x_1 + 3x_2 &= 0 \\ 2x_1 + 6x_2 &= 0 \end{aligned} \tag{12}$$

The line $x_1 + 3x_2 = 0$ is the null space of $\mathbf{A}$. For instance, $(-3, 1)$ is a point on this line. So are all scalar multiples of $(-3, 1)$. As you might have guessed, if the matrix has n rows and n columns, but is of rank $r < n$, then the nullspace is the $n - r$ dimensional subspace within $\mathbf{R^n}$. In this example, we are in $\mathbf{R^2}$, the 2-dimensional plane and the matrix is of rank 1. Therefore, we get a line as the nullspace.

2.3. **Factorizing a rectangular matrix using Singular Value Decomposition.** Now we have all the ingredients ready to take on the highlight of Linear Algebra: the SVD.

When $\mathbf{A}$ is a rectangular $m \times n$ matrix of rank r, we cannot factorize it as $\mathbf{S^{-1} \Lambda S}$ as that requires $\mathbf{A}$ to be square and be of full-rank. Instead, we solve two related problems. We construct two square matrices: $\mathbf{AA^T}$ and $\mathbf{A^T A}$ from the original matrix and compute two sets of eigenvalues and eigenvectors. Since both these new, square matrices are symmetric, their eigenvectors can be chosen to be orthonormal.

Let's assume that $\mathbf{A}$ is rectangular with $m > n$ and that it has been factorized as below:

$$A = U \Sigma V^T \tag{13}$$

where U is an $m \times m$ orthonormal matrix, V is $n \times n$ orthonormal matrix, and Σ is $m \times n$ such it only has non-zero entries σ_{ii} and all other entries are 0 when $i \neq j$. Note that since U and V are orthogonal (orthonormal by choice), $U^T U = I$ and $V^T V = I$. The proof of the existence of the above factorization is beyond the scope of this course. Let's trust that it has been shown that every matrix can be factorized as above.

If we multiply A with V, we get

$$AV = U \Sigma V^T V = U \Sigma \tag{14}$$

Or, if we take one column vector v_1 from V and the corresponding column from U as u_1, we get $Av_1 = \sigma_1 u_1$. The reason for this is that $dot(v_i, v_j) = 0 \ \forall i \neq j$ and $dot(v_i, v_i) = 1$ by the definition of orthogonality. If we take the transpose all around, we get $A^T = V \Sigma^T U^T$ and similarly we can show that $A^T u_i = \sigma_i v_i$.

Consider $A^TA = (V\Sigma^TU^T)(U\Sigma V^T) = V\Sigma^T\Sigma V^T = VDV^T$, where D is a diagonal matrix of dimension $n \times n$. Note that A^TA is a square matrix and therefore the above form suggests that V represent the eigenvectors of A^TA and D represent the corresponding eigenvalues. Analogously, for AA^T. So, we can factorize the two product matrices as below:

$$\begin{aligned} A^TA &= VDV^T \\ AA^T &= UD'U^T \end{aligned} \tag{15}$$

where $\mathbf{V}$ in an $n \times n$ orthonormal matrix consisting of eigenvectors of $\mathbf{A^TA}$ and $\mathbf{U}$ is an $m \times m$ orthonormal matrix containing the eigenvectors of $\mathbf{AA^T}$. $\mathbf{D}$ is a diagonal $n \times n$ matrix containing the eigenvalues of $\mathbf{A^TA}$. Similarly, $\mathbf{D'}$ is an $m \times m$ matrix, containing the eigenvalues of $\mathbf{AA^T}$. It turns out the $\mathbf{D}$ and $\mathbf{D'}$ contain the same non-zero eigenvalues (a total of r of them, where r is the rank of $\mathbf{A}$. Note that $r \leq n < m$. Further, we can easily work out that D contain entries that are σ_i^2. In other words, the singular values in Σ are the square roots of the eigenvalues of the product matrices AA^T and A^TA.

Since $r \leq n < m$, we will need $n - r$ additional vectors in V to make it $n \times n$ and $m - r$ additional vectors in U to make it $m \times m$. These additional vectors come from the null-space of A^T and A, respectively. The corresponding eigenvalues in D and D' are zero for these orthogonal vectors coming from the null-space. By extension, the corresponding singular values are 0 in the $m \times m$ Σ matrix.

We now showed how the SVD is solving two related eigenvalue problems by constructing two square matrices from the original rectangular matrix and taking an eigen decomposition of the two matrices and relating them back to the original matrix.