1 Linear algebra

In this section we will discuss vectors and matrices. We denote the (i, j)th entry of a matrix A as A_{ij}, and the ith entry of a vector as v_i.

1.1 Vectors and vector operations

A vector is a one dimensional matrix, and it can be written as a column vector:

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

or a row vector:

$$\boldsymbol{v} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$$

1.1.1 Dot product

The dot product of two equal-length vectors $\boldsymbol{u} = (u_1, \dots, u_n)$ and $\boldsymbol{v} = (v_1, \dots, v_n)$ is

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^T \boldsymbol{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

An equivalent definition is:

$$\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\| \|\boldsymbol{v}\| \cos\theta$$

where θ is the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$. Two vectors are orthogonal if their dot product is zero (i.e. $\theta = 90°$ or $\pi/2$). In 2D space we also call orthogonal vectors perpendicular.

1.1.2 Norm

The ℓ_2 norm, or length, of a vector $(v_1, \dots, v_n)$ is just $\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$. The norm of a vector v is usually written as $||v||$.

A more general norm is p-norm, denoted as

$$\|v\|_p = (\sum_{i=1}^{n} |v_i|^p)^{\frac{1}{p}}$$

when we take $p = 2$, we have the ℓ_2.

1.1.3 Triangle inequality

For two vectors u and v, we have

$$||u + v|| \leq ||u|| + ||v||$$

and

$$||u - v|| \geq \Big| ||u|| - ||v|| \Big|$$

1.2 Matrix operations

1.2.1 Matrix addition

Matrix addition is defined for matrices of the same dimension. Matrices are added componentwise:

$$\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} + \begin{bmatrix}5 & 6\\ 7 & 8\end{bmatrix} = \begin{bmatrix}1+5 & 2+6\\ 3+7 & 4+8\end{bmatrix} = \begin{bmatrix}6 & 8\\ 10 & 12\end{bmatrix}$$

1.2.2 Matrix multiplication

Matrices can be multiplied like so:

$$\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} \cdot \begin{bmatrix}5 & 6\\ 7 & 8\end{bmatrix} = \begin{bmatrix}1\times 5+2\times 7 & 1\times 6+2\times 8\\ 3\times 5+4\times 7 & 3\times 6+4\times 8\end{bmatrix} = \begin{bmatrix}19 & 22\\ 43 & 50\end{bmatrix}$$

You can also multiply non-square matrices, but the dimensions have to match (i.e. the number of columns of the first matrix has to equal the number of rows of the second matrix).

$$\begin{bmatrix}1 & 2\\ 3 & 4\\ 5 & 6\end{bmatrix} \cdot \begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{bmatrix} = \begin{bmatrix}1\cdot 1+2\cdot 4 & 1\cdot 2+2\cdot 5 & 1\cdot 3+2\cdot 6\\ 3\cdot 1+4\cdot 4 & 3\cdot 2+4\cdot 5 & 3\cdot 3+4\cdot 6\\ 5\cdot 1+6\cdot 4 & 5\cdot 2+6\cdot 5 & 5\cdot 3+6\cdot 6\end{bmatrix} = \begin{bmatrix}9 & 12 & 15\\ 19 & 26 & 33\\ 29 & 40 & 51\end{bmatrix}$$

In general, if matrix A is multiplied by matrix B, we have $(AB)_{ij} = \sum_k A_{ik}B_{kj}$ for all entries (i, j) of the matrix product.

Matrix multiplication is associative, i.e. $(AB)C = A(BC)$. It is also distributive, i.e. $A(B + C) = AB + AC$. However, it is **not** commutative. That is, AB does not have to equal BA.

Note that if you multiply a 1-by-n matrix with an n-by-1 matrix, that is the same as taking the dot product of the corresponding vectors.

1.2.3 Matrix transpose

The transpose operation switches a matrix's rows with its columns, so

$$\begin{bmatrix}1 & 2\\ 3 & 4\\ 5 & 6\end{bmatrix}^T = \begin{bmatrix}1 & 3 & 5\\ 2 & 4 & 6\end{bmatrix}$$

In other words, we define A^T by $(A^T)_{ij} = A_{ji}$.

Properties:

- $(A^T)^T = A$

- $(AB)^T = B^T A^T$

  Proof. Let $AB = C, (AB)^T = D$, then

$$\begin{aligned}
(AB)^T_{ij} &= D_{ij} = C_{ji} \\
&= \sum_k A_{jk} B_{ki} \\
&= \sum_k (A^T)_{kj} (B^T)_{ik} \\
&= \sum_k (B^T)_{ik} (A^T)_{kj} \\
\therefore D &= B^T A^T
\end{aligned}$$

□

- $(A + B)^T = A^T + B^T$

1.2.4 Identity matrix

The identity matrix I_n is an n-by-n matrix with all 1's on the diagonal, and 0's everywhere else. It is usually abbreviated I, when it is clear what the dimensions of the matrix are.

It has the property that when you multiply it by any other matrix, you get that matrix. In other words, if A is an m-by-n matrix, then $AI_n = I_m A = A$.

1.2.5 Matrix inverse

The inverse of a matrix A is the matrix that you can multiply A by to get the identity matrix. Not all matrices have an inverse. (The ones that have an inverse are called *invertible.*)

In other words, A^{-1} is the matrix where $AA^{-1} = A^{-1}A = I$ (if it exists).

Inverse matrix is unique for a particular matrix A, if the inverse exists.

Properties:

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1} A^{-1}$
- $(A^{-1})^T = (A^T)^{-1}$

  Proof.

$$(A^{-1})^T (A^T) = (AA^{-1})^T = I$$

□

1.3 Types of matrices

1.3.1 Diagonal matrix

A diagonal matrix is a matrix that has 0's everywhere except the diagonal. A diagonal matrix can be written $D = diag(d_1, d_2, \ldots, d_n)$, which corresponds to the matrix

$$D = \begin{bmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n \end{bmatrix}$$

You may verify that

$$D^k = \begin{bmatrix} d_1^k & 0 & \ldots & 0 \\ 0 & d_2^k & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n^k \end{bmatrix}$$

1.3.2 Triangular matrix

A lower triangular matrix is a matrix that has all its nonzero elements on or below the diagonal. An upper triangular matrix is a matrix that has all its nonzero elements on or above the diagonal.

1.3.3 Symmetric matrix

A is symmetric if $A = A^T$, i.e. $A_{ij} = A_{ji}$ for all entries (i, j) in A. Note that a matrix must be square in order to be symmetric.

1.3.4 Orthogonal matrix

A matrix U is orthogonal if $UU^T = U^T U = I$. (That is, the inverse of an orthogonal matrix is its transpose.)

Orthogonal matrices have the property that every row is orthogonal to every other row. That is, the dot product of any row vector with any other row vector is 0. In addition, every row is a unit vector, i.e. it has norm 1. (Try verifying this for yourself!)

Similarly, every column is a unit vector, and every column is orthogonal to every other column. (You can verify this by noting that if U is orthogonal, then U^T is also orthogonal.)

1.4 Linear independence and span

A linear combination of the vectors $v_1, \ldots, v_n$ is an expression of the form $a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$, where $a_1, \ldots, a_n$ are real numbers. Note that some of the a_i's may be zero.

The span of a set of vectors is the set of all possible linear combinations of that set of vectors.

The vectors $v_1, \ldots, v_n$ are linearly independent if you cannot find coefficients $a_1, \ldots, a_n$ where $a_1 v_1 + \cdots + a_n v_n = 0$ (except for the trivial solution $a_1 = a_2 = \cdots = 0$). Intuitively, this means you cannot write any of the vectors in terms of any linear combination of the other vectors. (A set of vectors is linearly dependent if it is not linearly independent.)

Important definitions and properties:

- A set of n independent n-vectors is called a **basis**.
- A set of $n+1$ (or more) n-vectors is always linearly dependent.

1.5 Eigenvalues and eigenvectors

Sometimes, multiplying a matrix by a vector just stretches that vector. If that happens, the vector is called an eigenvector of the matrix, and the "stretching factor" is called the eigenvalue.

Definition: Given a square matrix A, λ is an eigenvalue of A with the corresponding eigenvector x if $Ax = \lambda x$. (Note that in this definition, x is a vector, and λ is a number.)

(By convention, the zero vector cannot be an eigenvector of any matrix.)

Example: If

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

then the vector $\begin{bmatrix} 3 \\ -3 \end{bmatrix}$ is an eigenvector with eigenvalue 1, because

$$Ax = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \end{bmatrix} = 1 \cdot \begin{bmatrix} 3 \\ -3 \end{bmatrix}$$

1.5.1 Solving for eigenvalues and eigenvectors

We exploit the fact that $Ax = \lambda x$ if and only if $(A - \lambda I)x = 0$. (Note that λI is the diagonal matrix where all the diagonal entries are λ, and all other entries are zero.)

This equation has a nonzero solution for x if and only if the determinant of $A - \lambda I$ equals 0. (We won't prove this here, but you can google for "invertible matrix theorem".) Therefore, you can find the eigenvalues of the matrix A by solving the equation $det(A - \lambda I) = 0$ for λ. Once you have done that, you can find the corresponding eigenvector for each eigenvalue λ by solving the system of equations $(A - \lambda I)x = 0$ for x.

Example: If

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

then

$$A - \lambda I = \begin{bmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{bmatrix}$$

and

$$det(A - \lambda I) = (2 - \lambda)^2 - 1 = \lambda^2 - 4\lambda + 3$$

Setting this equal to 0, we find that $\lambda = 1$ and $\lambda = 3$ are possible eigenvalues.

To find the eigenvectors for $\lambda = 1$, we plug λ into the equation $(A - \lambda I)x = 0$. This gives us

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Any vector where $x_2 = -x_1$ is a solution to this equation, and in particular, $\begin{bmatrix} 3 \\ -3 \end{bmatrix}$ is one solution.

To find the eigenvectors for $\lambda = 3$, we again plug λ into the equation, and this time we get

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Any vector where $x_2 = x_1$ is a solution to this equation.

(**Note:** The above method is never used to calculate eigenvalues and eigenvectors for large matrices in practice, iterative methods are used instead.)

1.5.2 Properties of eigenvalues and eigenvectors

- Usually eigenvectors are normalized to unit length
- If A is symmetric, then all its eigenvalues are real
- The eigenvalues of any triangular matrix are its diagonal entries
- The trace of a matrix (i.e. the sum of the elements on its diagonal) is equal to the sum of its eigenvalues
- The determinant of a matrix is equal to the product of its eigenvalues

1.6 Matrix eigendecomposition

Theorem: Suppose A is an n-by-n matrix with n linearly independent eigenvectors. Then A can be written as $A = PDP^{-1}$, where P is the matrix whose columns are the eigenvectors of A, and D is the diagonal matrix whose entries are the corresponding eigenvalues.

In addition, $A^2 = (PDP^{-1})(PDP^{-1}) = PD^2P^{-1}$, and $A^n = PD^nP^{-1}$. (This is interesting because it's much easier to raise a diagonal matrix to a power than to exponentiate an ordinary matrix.)