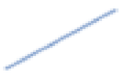
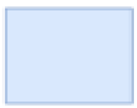
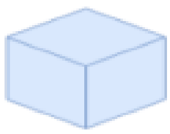
Basic Mathematics related to statistics

Linear Algebra

- Scalars, Vectors, Matrices and Tensors
- Multiplying Matrices and Vectors
- Identity and Inverse Matrices
- Linear Dependence and Span
- Norms

- Scalars: A scalar is just a single number, in contrast to most of the other objects studied in linear algebra, which are usually arrays of multiple numbers.
- Vectors: A vector is an array of numbers. The numbers are arranged in order. We can identify each individual number by its index in that ordering. Typically we give vectors lowercase names in bold typeface, such as x.
- Matrices: A matrix is a 2-D array of numbers, so each element is identified by two indices instead of just one.
- Tensors: In some cases we will need an array with more than two axes. In the general case, an array of numbers arranged on a regular grid with a variable number of axes is known as a tensor. We denote a tensor named "A" with this typeface: A. We

Type	Scalar	Vector	Matrix	Tensor
Definition	a single number	an array of numbers	2-D array of numbers	k-D array of numbers
Notation	x	$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$	$\boldsymbol{X} = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,n} \\ X_{2,1} & X_{2,2} & \dots & X_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ X_{m,1} & X_{m,2} & \dots & X_{m,n} \end{bmatrix}$	$\mathbf{X}$ $X_{i,j,k}$
Example	1.333	$\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ 9 \end{bmatrix}$	$\boldsymbol{X} = \begin{bmatrix} 1 & 2 & \dots & 4 \\ 5 & 6 & \dots & 8 \\ \vdots & \vdots & \vdots & \vdots \\ 13 & 14 & \dots & 16 \end{bmatrix}$	$x = \left[\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \\ 70 & 80 & 90 \end{bmatrix} \begin{bmatrix} 100 & 200 & 300 \\ 400 & 500 & 600 \\ 700 & 800 & 900 \end{bmatrix} \right]$
Python code example	x = np.array(1.333)	x = np.array([1,2,3, 4,5,6, 7,8,9])	x = np.array([[1,2,3,4], [5,6,7,8], [9,10,11,12], [13,14,15,16]])	x = np.array([[[1, 2, 3], [4, 5, 6], [7, 8, 9]], [[10, 20, 30], [40, 50, 60], [70, 80, 90]], [[100, 200, 300], [400, 500, 600], [700, 800, 900]]])
Visualization				 3-D Tensor

Operators

- A scalar can be thought of as a matrix with only a single entry. From this, we can see that a scalar is its own transpose: $a = a^T$.
- We can add matrices to each other, as long as they have the same shape, just by adding their corresponding elements: C = A + B where Ci,j = Ai,j + Bi,j.
- We can also add a scalar to a matrix or multiply a matrix by a scalar, just by performing that operation on each element of a matrix: D = a · B + c where Di,j =a·Bi,j +c.

- If A is of shape $m \times n$ and B is of shape $n \times p$, then C is of shape $m \times p$. We can write the matrix product just by placing two or more matrices together, for example, C = AB.
- Note that the standard product of two matrices is not just a matrix containing the product of the individual elements. Such an operation exists and is called the element-wise product, or Hadamard product, and is denoted as $A \odot B$.
- Singular:A square matrix with linearly dependent columns is known as singular.

- Norm:Measure the size of vectors using a function called a norm. Formally, the L^p norm
- L^2 Norm: With p = 2, is known as the Euclidean norm, which is simply the Euclidean distance from the origin to the point identified by x.

$$L^p : \|\boldsymbol{x}\|_p = \left(\sum_i |\, x_i \,|^p\right)^{\frac{1}{p}}, p \in \mathbb{R}, p \geq 1$$

$$L^1 : \|\boldsymbol{v}\|_1 = \left\| \begin{bmatrix} -1 \\ -2 \\ 3 \\ 4 \end{bmatrix} \right\|_1 = \sum_i |\, v_i \,| = |-1| + |-2| + |\,3\,| + |\,4\,| = 10$$

$$L^2 : \|\boldsymbol{v}\|_2 = \left\| \begin{bmatrix} -1 \\ -2 \\ 3 \\ 4 \end{bmatrix} \right\|_2 = \sqrt{\sum_i |\, v_i \,|^2} = \sqrt{|-1|^2 + |-2|^2 + |\,3\,|^2 + |\,4\,|^2} = \sqrt{30}$$

$$\|\boldsymbol{x}\|_\infty = max \,|\, x_i \,|$$

$$\|\boldsymbol{v}\|_\infty = \left\| \begin{bmatrix} -1 \\ -2 \\ 3 \\ 4 \end{bmatrix} \right\|_\infty = max \,|\, v_i \,| = |\,4\,| = 4$$

Sum of sq. diff.

$MSE(v_1, v_2) = \frac{1}{n} \sum_i (x_{1i} - x_{2i})^2$

Sum of absolute diff.

v_1	v_2
23	25
32	31
21	19
45	47
30	28
56	52
40	35

diff, abs, sum

$= |23-25| + |32-31| + |21-19| + |45-47|$
$+ |30-28| + |56-52| + |40-35|$

SAD $= 2+1+2+2+2+4+5 = 18$

$MAD = \frac{18}{7} = 2.57$

- Orthogonal: A vector x and a vector y are orthogonal to each other if $x^T y = 0$. If both vectors have nonzero norm, this means that they are at a 90 degree angle to each other. In R^n, at most n vectors may be mutually orthogonal with nonzero norm.
- Orthonormal: If the vectors not only are orthogonal but also have unit norm, we call them orthonormal.
- Orthogonal Matrix: An orthogonal matrix is a square matrix whose rows are mutually orthonormal and whose columns are mutually orthonormal.
  $A^T A = AA^T = I$
  $A^{-1} = A^T$

$$\boldsymbol{u}^\intercal \boldsymbol{v} = 0 \text{ and } \|\boldsymbol{u}\| = 1, \|\boldsymbol{v}\| = 1$$

$$\boldsymbol{u}^\intercal \boldsymbol{v} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{2}}{2} \\ \frac{1}{2} \end{bmatrix} = (\tfrac{1}{\sqrt{2}} \times \tfrac{1}{2}) + (0 \times \tfrac{\sqrt{2}}{2}) + (-\tfrac{1}{\sqrt{2}} \times \tfrac{1}{2}) = 0$$

$$\sqrt{{u_1}^2 + {u_2}^2 + {u_3}^2} = \sqrt{(\tfrac{1}{\sqrt{2}})^2 + 0^2 + (-\tfrac{1}{\sqrt{2}})^2} = 1$$

$$\sqrt{{v_1}^2 + {v_2}^2 + {v_3}^2} = \sqrt{(\tfrac{1}{2})^2 + (\tfrac{\sqrt{2}}{2})^2 + (\tfrac{1}{2})^2} = 1$$

Therefore, we say that, vector $\boldsymbol{u}$ and vector $\boldsymbol{v}$ are orthonormal.

Orthogonal Matrix Properties

- All identity matrices are an orthogonal matrix.
- The product of two orthogonal matrices is also an orthogonal matrix.
- The collection of the orthogonal matrix of order n × n, in a group, is called an orthogonal group and is denoted by 'O'.
- The transpose of the orthogonal matrix is also orthogonal.
- The inverse of the orthogonal matrix, which is A-1 is also an orthogonal matrix.
- The determinant of the orthogonal matrix has a value of 1.
- The eigenvalues of the orthogonal matrix also have a value as 1, and its eigenvectors would also be orthogonal and real.

Example 1:

$$u^T = \begin{bmatrix} \frac{4}{\sqrt{100}} & \frac{9}{\sqrt{100}} & \frac{\sqrt{2}}{\sqrt{100}} & \frac{-1}{\sqrt{100}} \end{bmatrix}$$

$$v = \begin{bmatrix} \frac{9}{\sqrt{100}} \\ \frac{-4}{\sqrt{100}} \\ \frac{1}{\sqrt{100}} \\ \frac{\sqrt{2}}{\sqrt{100}} \end{bmatrix}$$

if $\|u\|_2 = 1$ then u is orthonormal.

if $\|v\|_2 = 1$ then v is orthonormal.

$$\|u\|_2 = \frac{16 + 81 + 2 + 1}{100} = \frac{100}{100} = 1$$

$$\|v\|_2 = \frac{81 + 16 + 1 + 2}{100} = \frac{100}{100} = 1$$

L_2 norm or product with its own then orthonormal

Example 2:

$$u^T = \begin{bmatrix} \frac{4}{\sqrt{100}} & \frac{9}{\sqrt{100}} & \frac{\sqrt{2}}{\sqrt{100}} & \frac{-1}{\sqrt{100}} \end{bmatrix}$$

$$v = \begin{bmatrix} 9/\sqrt{100} \\ -4/\sqrt{100} \\ 1/\sqrt{100} \\ \sqrt{2}/\sqrt{100} \end{bmatrix}$$

if $u^T v = 0$, then u & v are orthogonal

$$\frac{36}{100} + \frac{-36}{100} + \frac{\sqrt{2}}{100} - \frac{\sqrt{2}}{100}$$

$$u^T v = 0$$

Example 3:

$$A = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$AA^T = I$$

$$A^T = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$AA^T = \begin{bmatrix} \cos^2\theta + \sin^2\theta & & -\sin\theta\cos\theta + \sin\theta\cos\theta \\ & & 0 \\ -\sin\theta\cos\theta + \cos\theta\sin\theta & 0 & \sin^2\theta + \cos^2\theta \end{bmatrix}$$

Eigen Decomposition

- Decompose a matrix into a set of eigenvectors and eigenvalues.
- An eigenvector of a square matrix A is a nonzero vector v such that multiplication by A alters only the scale of v: $Av = \lambda v$.
- The scalar λ is known as the eigenvalue corresponding to this eigenvector. (One can also find a left eigenvector such that $v^T A = \lambda v^T$, but we are usually concerned with right eigenvectors.)
- The eigendecomposition of A is then given by $A = V diag(\lambda) V^1$.

The formal statement of the problem

For a given square matrix, A, we investigate the existence of column vectors, X, such that

$$\mathrm{AX} = \lambda \mathrm{X},$$

for some scalar quantity λ.

Each such column vector is called an **"eigenvector"** of the matrix, A.

Each corresponding value of λ is called an **"eigenvalue"** of the matrix, A.

Notes:

(i) The German word "eigen" means"hidden".

(ii) Other alternative names are "latent values and latent vectors" or "characteristic values and characteristic vectors".

(iii) In the discussion which follows, A will be, mostly, a matrix of order 3×3.

Let

$$\mathrm{A} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \quad \text{and} \quad \mathrm{X} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Then the matrix equation, $\mathrm{AX} = \lambda \mathrm{X}$, means

$$\begin{aligned} a_1x + b_1y + c_1z &= \lambda x, \\ a_2x + b_2y + c_2z &= \lambda y, \\ a_3x + b_3y + c_3z &= \lambda z; \end{aligned}$$

or, on rearrangement,

$$\begin{aligned} (a_1 - \lambda)x + b_1y + c_1z &= 0, \\ a_2x + (b_2 - \lambda)y + c_2z &= 0, \\ a_3x + b_3y + (c_3 - \lambda)z &= 0. \end{aligned}$$

This is a set of homogeneous linear equations in x, y and z and may be written

$$(\mathrm{A} - \lambda \mathrm{I})\mathrm{X} = [0],$$

where I denotes the identity matrix of order 3×3.

$$|\mathrm{A} - \lambda\mathrm{I}| = \begin{vmatrix} a_1 - \lambda & b_1 & c_1 \\ a_2 & b_2 - \lambda & c_2 \\ a_3 & b_3 & c_3 - \lambda \end{vmatrix} = 0.$$

On expansion, this gives a cubic equation in λ called the **"characteristic equation"** of A.

The left-hand side of the characteristic equation is called the **"characteristic polynomial"** of A.

The characteristic equation of a 3×3 matrix, being a <u>cubic</u> equation, will (in general) have three solutions.

1. Determine the eigenvalues and eigenvectors of the matrix

$$\mathrm{A} = \begin{bmatrix} 2 & 4 \\ 5 & 3 \end{bmatrix}.$$

Solution

(a) The eigenvalues

The characteristic equation is given by

$$0 = \begin{vmatrix} 2 - \lambda & 4 \\ 5 & 3 - \lambda \end{vmatrix} = \lambda^2 - 5\lambda - 14 = (\lambda + 2)(\lambda - 7).$$

The eigenvalues are therefore $\lambda = -2$ and $\lambda = 7$.

(b) The eigenvectors

Case 1. $\lambda = -2$

We require to solve the equation $x + y = 0$,

giving $x : y = -1 : 1$ and a corresponding eigenvector

$$\mathrm{X} = \alpha \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

where α is any **non-zero** scalar.

Case 2. $\lambda = 7$

We require to solve the equation $5x - 4y = 0$,

giving $x : y = 4 : 5$ and a corresponding eigenvector

$$\mathrm{X} = \beta \begin{bmatrix} 4 \\ 5 \end{bmatrix},$$

where β is any **non-zero** scalar.

(b) The eigenvectors

Case 1. $\lambda = -2$

We require to solve the equation $x + y = 0$,

giving $x : y = -1 : 1$ and a corresponding eigenvector

$$X = \alpha \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

where α is any **non-zero** scalar.

Case 2. $\lambda = 7$

We require to solve the equation $5x - 4y = 0$,

giving $x : y = 4 : 5$ and a corresponding eigenvector

$$X = \beta \begin{bmatrix} 4 \\ 5 \end{bmatrix},$$

where β is any **non-zero** scalar.

(i) The eigenvalues of a matrix are the same as those of its transpose.

Proof:

Given a square matrix, A, the eigenvalues of A^T are the solutions of the equation

$$|A^T - \lambda I| = 0.$$

But, since I is a symmetric matrix, this is equivalent to

$$|(A - \lambda I)^T| = 0.$$

The result follows, since a determinant is unchanged in value when it is transposed.

(ii) The Eigenvalues of the multiplicative inverse of a matrix are the reciprocals of the eigenvalues of the matrix itself.

Proof:

If λ is any eigenvalue of a square matrix, A, then

$$AX = \lambda X,$$

for some column vector, X.

Premultiplying this relationship by A^{-1}, we obtain

$$A^{-1}AX = A^{-1}(\lambda X) = \lambda(A^{-1}X).$$

Thus,

$$A^{-1}X = \frac{1}{\lambda}X.$$

(iii) The eigenvectors of a matrix and its multiplicative inverse are the same.

Proof:

This follows from the proof of **(ii)**, since

$$A^{-1}X = \frac{1}{\lambda}X$$

implies that X is an eigenvector of A^{-1}.

(iv) If a matrix is multiplied by a single number, the eigenvalues are multiplied by that number, but the eigenvectors remain the same.

Proof:

If A is multiplied by α, we may write the equation $AX = \lambda X$ in the form $\alpha AX = \alpha\lambda X$.

Thus, αA has eigenvalues, $\alpha\lambda$, and eigenvectors, X.

(v) If $\lambda_1, \quad \lambda_2, \quad \lambda_3, \ . \ .$ are the eigenvalues of the matrix A and n is a positive integer, then $\lambda_1^n, \quad \lambda_2^n, \quad \lambda_3^n, \ . \ .$ are the eigenvalues of A^n.

Proof:

If λ denotes any one of the eigenvalues of the matrix, A, then $AX = \lambda X$.

Premultiplying both sides by A, we obtain
$A^2X = A\lambda X = \lambda AX = \lambda^2 X$.

(vi) If $\lambda_1, \ \lambda_2, \ \lambda_3, \ \ . \ . \ .$ are the eigenvalues of the $n \times n$ matrix A, I is the $n \times n$ multiplicative identity matrix and k is a single number, then the eigenvalues of the matrix $A + kI$ are $\lambda_1 + k, \ \lambda_2 + k, \ \lambda_3 + k, \ . \ . \ . .$

Proof:

If λ is any eigenvalue of A, then $AX = \lambda X$.

Hence,

$$(A + kI)X = AX + kX = \lambda X + kX = (\lambda + k)X.$$

(vii) A matrix is singular ($|A| = 0$) if and only if at least one eigenvalue is equal to zero.

Proof:

(a) If X is an eigenvector corresponding to an eigenvalue, $\lambda = 0$, then $AX = \lambda X = [0]$.

From the theory of homogeneous linear equations, it follows that $|A| = 0$.

(b) Conversely, if $|A| = 0$, the homogeneous system $AX = [0]$ has a solution for X other than $X = [0]$.

Hence, at least one eigenvalue must be zero.

(viii) If A is an orthogonal matrix ($AA^T = I$), then every eigenvalue is either $+1$ or -1.

Proof:

The statement $AA^T = I$ can be written $A^{-1} = A^T$ so that, by **(i)** and **(ii)**, the eigenvalues of A are equal to their own reciprocals

That is, they must have values $+1$ or -1.

(ix) If the elements of a matrix below the leading diagonal or the elements above the leading diagonal are all equal zero, then the eigenvalues are equal to the diagonal elements.