Machine Learning: Algorithms and Applications

Philip O. Ogunbona

Advanced Multimedia Research Lab
University of Wollongong

Math Review - Linear Algebra
Autumn 2024

Outline of Topics

Introduction

- To provide context, imagine that you are a data scientist charged with the task of capturing some of the characteristics that describe a person. You may consider measuring height, weight, circumference of neck, circumference of waist and possibly age.
- These five quantities (variables) will vary from person to person. Actually they are random variables.
- Let us represent them by x_1, x_2, x_3, x_4, x_5 respectively.
- A convenient data structure to model these variables for each person we consider is a vector.
- When we consider a group of people we can use another data structure, matrix, to model the information about the group.

Linear algebra provides a mathematical framework to reason about and manipulate vectors and matrices. This is used extensively in Machine Learning to model and solve problems.

Vectors

With the context in mind we now consider elementary concepts in linear algebra.

- A d-dimensional column vector and its transpose (a row vector) can be written as,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \quad \text{and} \quad \mathbf{x}^t = \begin{bmatrix} x_1 & x_2 & \dots & x_d \end{bmatrix}$$

  We assume that all the components can take on real values. The transpose can also be written as $\mathbf{x}'$.

Vectors

With the context in mind we now consider elementary concepts in linear algebra.

- A d-dimensional column vector and its transpose (a row vector) can be written as,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \quad \text{and} \quad \mathbf{x}^t = \begin{bmatrix} x_1 & x_2 & \dots & x_d \end{bmatrix}$$

  We assume that all the components can take on real values. The transpose can also be written as $\mathbf{x}'$.

Example (Description of one person (5-dimensional vector))

$$\mathbf{x} = \begin{bmatrix} 1.5 \\ 75.2 \\ 41.3 \\ 81.28 \\ 35.5 \end{bmatrix} \quad \text{and} \quad \mathbf{x}^t = \begin{bmatrix} 1.5 & 75.2 & 41.3 & 81.28 & 35.5 \end{bmatrix}$$

Matrices

- An $n \times d$ matrix $\mathbf{M}$ and its $d \times n$ transpose $\mathbf{M}^t$ are written as,

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1d} \\ m_{21} & m_{22} & \dots & m_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \dots & m_{nd} \end{bmatrix} \; ; \quad \mathbf{M}^t = \begin{bmatrix} m_{11} & m_{21} & \dots & m_{n1} \\ m_{12} & m_{22} & \dots & m_{n2} \\ \vdots & \vdots & \ddots & \dots \\ m_{1d} & m_{2d} & \dots & m_{nd} \end{bmatrix}$$

Matrices

- An $n \times d$ matrix $\mathbf{M}$ and its $d \times n$ transpose $\mathbf{M}^t$ are written as,

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1d} \\ m_{21} & m_{22} & \dots & m_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \dots & m_{nd} \end{bmatrix} \; ; \; \mathbf{M}^t = \begin{bmatrix} m_{11} & m_{21} & \dots & m_{n1} \\ m_{12} & m_{22} & \dots & m_{n2} \\ \vdots & \vdots & \ddots & \dots \\ m_{1d} & m_{2d} & \dots & m_{nd} \end{bmatrix}$$

Example (Description of five variables for 3 people)

A 3×5 matrix $\mathbf{M}$ and its 5×3 transpose $\mathbf{M}^t$ are written as,

$$\mathbf{M} = \begin{bmatrix} 1.5 & 75.2 & 41.3 & 81.28 & 35.5 \\ 1.75 & 80.6 & 46.7 & 102.5 & 45 \\ 1.82 & 69.3 & 42.5 & 83.5 & 30 \end{bmatrix} \; ; \; \mathbf{M}^t = \begin{bmatrix} 1.5 & 1.75 & 1.82 \\ 75.2 & 80.6 & 69.3 \\ 41.3 & 46.7 & 42.5 \\ 81.28 & 102.5 & 83.5 \\ 35.5 & 45 & 30 \end{bmatrix}$$

Matrices - algebra

- We can write the entries of a matrix M as m_{ij} where i and j refer to the row and column values respectively.
- The product, **C**, of two matrices **A** (with dimension $m \times n$) and **B** (with dimension $n \times d$) is given as,

  $$\mathbf{C} = \mathbf{AB}$$

  and the entries of **C** are c_{ij}. Each entry is given by,

  $$c_{ij} = \sum_{k=1}^{j} a_{ik} \times b_{kj}$$

- We note that the transpose of the matrix C is,

  $$C^t = (AB)^t = B^t A^t$$

Matrices - algebra

Example (Product)

The product, **M**, of two **compatible** matrices **A** and **B** is given as,

$$\mathbf{M} = \begin{bmatrix} 4.6 & 5.7 & 6.1 & 5.5 \\ 2.4 & 3.6 & 4.7 & 4.9 \\ 3.5 & 5.3 & 9.5 & 8.5 \end{bmatrix} \times \begin{bmatrix} 3.5 & 6.2 & 1.0 \\ 1.5 & 3.7 & 3.3 \\ 4.1 & 8.7 & 7.5 \\ 7.5 & 4.1 & 9.5 \end{bmatrix} = \begin{bmatrix} 90.91 & 125.23 & 121.41 \\ 69.82 & 89.18 & 96.08 \\ 122.90 & 158.81 & 172.99 \end{bmatrix}$$

Notice that **A** is of dimension 3×4 while **B** is 4×3. The entry m_{11} of **M** is obtained as,

$$m_{11} = 4.6 \times 3.5 + 5.7 \times 1.5 + 6.1 \times 4.1 + 5.5 \times 7.5 = 90.91$$

Similarly, the entry m_{32},

$$m_{32} = 3.5 \times 6.2 + 5.3 \times 3.7 + 9.5 \times 8.7 + 8.5 \times 4.1 = 158.81$$

Matrices - algebra

- We can multiply a matrix, **M**, and a vector, **x**, to obtain a vector, **y**,

$$\begin{bmatrix} m_{11} & m_{12} & \dots & m_{1d} \\ m_{21} & m_{22} & \dots & m_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \dots & m_{nd} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} y_1 \\ x_y \\ \vdots \\ y_n \end{bmatrix}$$

  Each component of the vector, **y** is given by,

$$y_i = \sum_{j=1}^{d} m_{ij} x_j$$

- The number of columns of **M** must be equal to the number of rows of **x**

Matrices - algebra

Example

The product, **M**, of a matrix **A** and vector **x** is given as,

$$\mathbf{M} = \begin{bmatrix} 4.6 & 5.7 & 6.1 & 5.5 \\ 2.4 & 3.6 & 4.7 & 4.9 \\ 3.5 & 5.3 & 9.5 & 8.5 \end{bmatrix} \times \begin{bmatrix} 4.1 \\ 8.7 \\ 7.5 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 122.45 \\ 83.76 \\ 144.46 \end{bmatrix}$$

Matrices - algebra

Example

The product $\mathbf{A} \times \mathbf{B}$ is not always equal to $\mathbf{B} \times \mathbf{A}$. Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$. We have $\mathbf{AB} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$ while $\mathbf{BA} = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix}$

Matrices - algebra

Example

Notice that $\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \times \begin{bmatrix} 2 & 4 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ while $\begin{bmatrix} 2 & 4 \\ -1 & -2 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 14 & 28 \\ -7 & -14 \end{bmatrix}$

Matrices

Definition

A square matrix **M**, (with dimension $d \times d$) is called symmetric if the entries have the following relationship,

$$m_{ij} = m_{ji}$$

Definition

A square matrix **M**, (with dimension $d \times d$) is called skew-symmetric (or anti-symmetric) if the entries have the following relationship,

$$m_{ij} = -m_{ji}$$

Example

The matrix, $\mathbf{C} = \begin{bmatrix} 1 & 2 & 4 & 5 \\ 2 & -8 & 8 & 1 \\ 4 & 8 & 5 & 3 \\ 5 & 1 & 3 & 7 \end{bmatrix}$ is symmetric and $\mathbf{D} = \begin{bmatrix} 0 & -2 & 4 & -5 \\ 2 & 0 & 8 & 1 \\ -4 & -8 & 0 & 3 \\ 5 & -1 & -3 & 0 \end{bmatrix}$ is skew-symmetric.

Matrices

Definition

A general matrix **M**, is called non-negative if,

$$m_{ij} \geq 0, \text{ for all } i \text{ and } j$$

Example

The matrix, $\mathbf{B} = \begin{bmatrix} 1 & 2 & 0 & 5 \\ 2 & 8 & 8 & 1 \\ 4 & 8 & 5 & 3 \\ 5 & 0 & 3 & 7 \end{bmatrix}$ is non-negative.

Matrices

Definition

A square matrix $\mathbf{I}$, (with dimension $d \times d$) is the identity matrix and has the diagonal entries equal to unity (1) and other entries zero (0). The Kronecker delta function or Kronecker symbol, defined as

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{otherwise.} \end{cases}$$

can serve to define entries of an identity matrix.

Example

The matrix, $\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ is identity of dimension 4.

Matrices

Definition

A general diagonal matrix is one that has a zero (0) in all the off-diagonal entries and denoted as $diag(m_{11}, m_{22}, \ldots, m_{dd})$

Example

The matrix $\mathbf{D} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ is diagonal. It can be written as $\mathrm{diag}(4, 6, 2, 1)$.

Matrices

Definition

Addition of vectors and of matrices is component by component.

Example (Sum of matrices)

The sum, **M**, of two **compatible** matrices **A** and **B** is given as,

$$\mathbf{M} = \begin{bmatrix} 4.6 & 5.7 & 6.1 \\ 2.4 & 3.6 & 4.7 \\ 3.5 & 5.3 & 9.5 \end{bmatrix} + \begin{bmatrix} 3.5 & 6.2 & 1.0 \\ 1.5 & 3.7 & 3.3 \\ 4.1 & 8.7 & 7.5 \end{bmatrix} = \begin{bmatrix} 8.1 & 11.9 & 7.1 \\ 3.9 & 7.3 & 8.0 \\ 7.6 & 14.0 & 17.0 \end{bmatrix}$$

Inner product

Definition

The inner product (or scalar product) of two vectors, $\mathbf{x}$ and $\mathbf{y}$ having the same dimensionality, d, will be denoted as $\mathbf{x}^t\mathbf{y}$ and the result is a scalar,

$$\mathbf{x}^t\mathbf{y} = \sum_{i=1}^{d} \mathbf{x}_i\mathbf{y}_i = \mathbf{y}^t\mathbf{x}$$

Definition

The Euclidean norm or length of a vector $\mathbf{x}$ is,

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^t\mathbf{x}}$$

The vector is normalized if $\|\mathbf{x}\| = 1$.

Inner product

Definition

The inner product (or scalar product) of two vectors, $\mathbf{x}$ and $\mathbf{y}$ having the same dimensionality, d, will be denoted as $\mathbf{x}^t\mathbf{y}$ and the result is a scalar,

$$\mathbf{x}^t\mathbf{y} = \sum_{i=1}^{d} \mathbf{x}_i \mathbf{y}_i = \mathbf{y}^t\mathbf{x}$$

Definition

The angle, θ, between two d-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$ is given by,

$$cos\theta = \frac{\mathbf{x}^t\mathbf{y}}{||\mathbf{x}||\,||\mathbf{y}||}$$

Inner product is a measure of the colinearity of two vectors; an indication of similarity (to within a scale factor).

Inner product

Example (Inner product)

The inner product of two vectors $\mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 6 \\ 8 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 5 \\ 1 \\ 0 \\ 2 \\ 7 \end{bmatrix}$ is

$1 \times 5 + 3 \times 1 + 4 \times 0 + 6 \times 2 + 8 \times 7 = 76$

Inner product

Example (Inner product)

The inner product of two vectors $\mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 6 \\ 8 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 5 \\ 1 \\ 0 \\ 2 \\ 7 \end{bmatrix}$ is

$1 \times 5 + 3 \times 1 + 4 \times 0 + 6 \times 2 + 8 \times 7 = 76$

Example (Magnitude)

The magnitude of vector $\mathbf{x}$ is $\|\mathbf{x}\| = \sqrt{1 \times 1 + 3 \times 3 + 4 \times 4 + 6 \times 6 + 8 \times 8} = \sqrt{126} = 11.23$.

Inner product

Example (Inner product)

The inner product of two vectors $\mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 6 \\ 8 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 5 \\ 1 \\ 0 \\ 2 \\ 7 \end{bmatrix}$ is

$1 \times 5 + 3 \times 1 + 4 \times 0 + 6 \times 2 + 8 \times 7 = 76$

Example (Magnitude)

The magnitude of vector $\mathbf{x}$ is $||\mathbf{x}|| = \sqrt{1 \times 1 + 3 \times 3 + 4 \times 4 + 6 \times 6 + 8 \times 8} = \sqrt{126} = 11.23$.

Example (Magnitude)

The magnitude of vector $\mathbf{y}$ is $||\mathbf{y}|| = \sqrt{5 \times 5 + 1 \times 1 + 0 + 2 \times 2 + 7 \times 7} = \sqrt{79} = 8.89$.

Inner product

Example (Inner product)

The inner product of two vectors $\mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ 4 \\ 6 \\ 8 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 5 \\ 1 \\ 0 \\ 2 \\ 7 \end{bmatrix}$ is

$1 \times 5 + 3 \times 1 + 4 \times 0 + 6 \times 2 + 8 \times 7 = 76$

Example (Angle)

The angle between the vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\theta = \arccos \frac{\mathbf{x}^t \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} = \arccos \frac{76}{11.23 \times 8.89} = 0.707 \text{ radians}$$

Orthogonal vectors

Definition

If $\mathbf{x}^t\mathbf{y} = 0$, the vectors are orthogonal.

Definition

If $\mathbf{x}^t\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|$, the vectors are colinear.

Definition

The Cauchy-Schwartz inequality follows directly from previous definition of an angle between two vectors. In other words,

$$\mathbf{x}^t\mathbf{y} \leq \|\mathbf{x}\|\|\mathbf{y}\|.$$

Matrices

Definition

The trace of a square matrix **A**, denoted $Tr\{\mathbf{A}\}$, is the sum of its diagonal elements,

$$Tr\{A\} = \sum_{i=1}^{d} a_{ii}$$

and $Tr\{\mathbf{CD}\} = Tr\{\mathbf{DC}\}$ provided that the product **CD** is a square matrix. Neither **C** nor **D** need be square.

Example

The trace of the matrix $\mathbf{D} = \begin{bmatrix} 4 & 2 & 9 & 0 \\ 1 & 6 & 7 & 1 \\ 4 & 9 & 2 & 3 \\ 2 & 7 & 3 & 1 \end{bmatrix}$ is $4 + 6 + 2 + 1 = 13$.

Matrices

Definition

The determinant of a square $(d \times d)$ matrix $\mathbf{M}$, written as $|\mathbf{M}|$ is the sum,

$$|\mathbf{M}| = \sum_{j=1}^{d} m_{ij} M_{ij} \text{ for } i = 1 \ldots, d$$

where the cofactor, M_{ij}, is the determinant of the matrix formed by deleting the ith row and the jth column of $\mathbf{M}$, multiplied by $(-1)^{i+j}$.

Definition

The transpose of the matrix of the cofactors, $\mathbf{C}(c_{ij} = M_{ij})$, is called the adjoint of $\mathbf{M}$, Adj[$\mathbf{M}$].

Matrices

Definition

The inverse of a $d \times d$ matrix, $\mathbf{M}$ is that unique matrix $\mathbf{M}^{-1}$ of dimension $d \times d$, with entries such that,

$$\mathbf{M}^{-1}\mathbf{M} = \mathbf{M}\mathbf{M}^{-1} = \mathbf{I}$$

Definition

We can obtain the inverse of a matrix from,

$$\mathbf{M}^{-1} = \frac{Adj[\mathbf{M}]}{|\mathbf{M}|}$$

Definition

If the inverse exists the matrix is said to be nonsingular otherwise it is singular and $|\mathbf{M}| = 0$

Note that, $(\mathbf{M}^t)^{-1} = (\mathbf{M}^{-1})^t$ and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Matrices

Example

Consider the matrix $\mathbf{A} = \begin{bmatrix} 2 & 4 & 6 \\ -1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix}$. The determinant is,

$$2 \times (2 \times 9 - 3 \times 4) - 4 \times (-1 \times 9 - 1 \times 3) + 6 \times (-1 \times 4 - 2 \times 1) = 24$$

Example

The adjoint of $\mathbf{A}$ is $\begin{bmatrix} 6 & 12 & -6 \\ -12 & 12 & -4 \\ 0 & -12 & 8 \end{bmatrix}^t = \begin{bmatrix} 6 & -12 & 0 \\ 12 & 12 & -12 \\ -6 & -4 & 8 \end{bmatrix}$.

Example

The inverse of $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \frac{1}{24} \begin{bmatrix} 6 & -12 & 0 \\ 12 & 12 & -12 \\ -6 & -4 & 8 \end{bmatrix}.$$

Linear independence and rank

Definition

A set of k vectors of equal dimension are linearly dependent if there exists a set of scalars $c_1, c_2, \ldots, c_k$, not all zero, such that,

$$c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k = 0$$

If it is impossible to find such a set of $c_1, c_2, \ldots, c_k$, then the vectors, $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are said to be linearly independent.

Definition

The rank of a matrix is the maximum number of linearly independent rows (or equivalently, the maximum number of linearly independent columns).

Definition

A $d \times d$ matrix is of full rank if the rank is equal to d. It will also be true that the determinant is non-zero and it will possess an inverse.

Linear independence and rank

Definition

For a rectangular matrix (i.e. non-square) $\mathbf{M}$ of dimension $d \times n$, the rank of $\mathbf{M}$, denoted rank $(\mathbf{M})$ is such that rank $(\mathbf{M}) \leq min(d, n)$.

We have that,

$$rank(\mathbf{M}) = rank(\mathbf{M^t}) = rank(\mathbf{M^tM}) = rank(\mathbf{MM^t})$$

Orthogonal matrix

Definition

A square matrix $\mathbf{M}$, is orthogonal if,

$$\mathbf{M}\mathbf{M}^t = \mathbf{M}^t\mathbf{M} = \mathbf{I}$$

- The rows and columns are orthonormal, $\mathbf{x}^t\mathbf{y} = 0$ and $\mathbf{x}^t\mathbf{x} = 1$, $\mathbf{y}^t\mathbf{y} = 1$ for any two different columns $\mathbf{x}$ and $\mathbf{y}$.
- An orthogonal matrix represents a linear transformation that preserves distances and angles, consisting of a rotation and/or reflection
- An orthogonal matrix is nonsingular and the inverse is its transpose, $\mathbf{M}^{-1} = \mathbf{M}^t$
- The determinant of an orthogonal matrix is ± 1, with -1 indicating a reflection and $+1$ indicating pure rotation.

Positive definiteness

Definition

A square matrix $\mathbf{M}$ is positive definite if the quadratic form, $\mathbf{x}^{\mathbf{t}}\mathbf{M}\mathbf{x} > \mathbf{0}$ for all vectors $\mathbf{x} \neq \mathbf{0}$.

Definition

A square matrix $\mathbf{M}$ is positive semidefinite if the quadratic form, $\mathbf{x}^{\mathbf{t}}\mathbf{M}\mathbf{x} \geq \mathbf{0}$ for all vectors $\mathbf{x} \neq \mathbf{0}$. A positive definite matrix will have a full rank.

Eigenvalue problem

- Given a $d \times d$ matrix $\mathbf{M}$, an important class of linear equations is of the form,

  $$\mathbf{Mx} = \lambda\mathbf{x}$$

  or

  $$(\mathbf{M} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$$

  for a scalar λ.
- The solution to the characteristic equation,

  $$|M - \lambda I| = 0$$

  gives the eigenvalues or characteristic roots of the $d \times d$ matrix.

Eigenvalue problem

- The characteristic equation is a dth order polynomial in λ. There are d solutions, $\lambda_1, \lambda_2, \ldots, \lambda_d$. They are not necessarily distinct and may be real or complex.
- Associated with each eigenvalue, λ_i is an eigenvector, $\mathbf{u}_i$, such that,

$$\mathbf{M}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

- The eigenvectors are not unique because any scalar multiple of $\mathbf{u}_i$ satisfies $\mathbf{M}\mathbf{u}_i = \lambda_i \mathbf{u}_i$. Eigenvectors are usually normalized so that, $\mathbf{u}_i^t \mathbf{u}_i = 1$.

Eigenvalue problem

Properties of eigenvalues and eigenvectors

1. The product of the eigenvalues is equal to the determinant of **M**. $\prod_{i=1}^{d} \lambda_i = \det(\mathbf{M})$. For a given matrix, if the eigenvalues are all non-zero, then the inverse of **M** exists.
2. The sum of the eigenvalues is equal to the trace of the matrix. $\sum_{i=1}^{d} \lambda_i = Tr(\mathbf{M})$.
3. If **M** is a real symmetric matrix, the eigenvalues and eigenvectors are all real.
4. If **M** is positive definite, the eigenvalues are all greater than zero.
5. If **M** is positive semidefinite of rank m, then there will be m non-zero eigenvalues and $d - m$ eigenvalues with the value of zero.

Eigenvalue problem

Properties of eigenvalues and eigenvectors (continued)

6. Every real symmetric matrix has a set of orthonormal characteristic vectors. The matrix, **U**, whose columns are the eigenvectors of the real symmetric matrix is orthogonal. $\mathbf{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_d\}$. We have $\mathbf{U}^t\mathbf{U} = \mathbf{U}\mathbf{U}^t = \mathbf{I}$.

7. The matrix **U** diagonalizes **M**,

$$\mathbf{U}^t\mathbf{M}\mathbf{U} = \Lambda$$

where $\Lambda = diag(\lambda_1, \ldots, \lambda_d)$ is a diagonal matrix whose entries are the eigenvalues of **M**.

$$\mathbf{M} = \mathbf{U}\Lambda\mathbf{U}^t = \sum_{i=1}^{d} \lambda_i \mathbf{u}_i \mathbf{u}_i^t$$

Eigenvalue problem

Properties of eigenvalues and eigenvectors (continued)

8. If **M** is positive definite, then $\mathbf{M}^{-1} = \mathbf{U}\Lambda^{-1}\mathbf{U}^t$. Here, $\Lambda^{-1} = diag(1/\lambda_1, \ldots, 1/\lambda_d)$

Singular value decomposition

SVD - singular value decomposition

The compact singular value decomposition of matrix $\boldsymbol{M} \in \mathbb{R}^{m \times n}$ with $r = \operatorname{rank}(\boldsymbol{M}) \leq \min(m, n)$ can be written as follows:

$$\boldsymbol{M} = \boldsymbol{U}_M \boldsymbol{\Sigma}_M \boldsymbol{V}_M^t.$$

- The $r \times r$ matrix, $\boldsymbol{\Sigma}_M = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$, is diagonal and contains the non-zero singular values of $\boldsymbol{M}$ sorted in decreasing order, that is $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0$.
- The matrices $\boldsymbol{U}_M \in \mathbb{R}^{m \times r}$ and $\boldsymbol{V}_M \in \mathbb{R}^{n \times r}$ have orthonormal columns that contain the left and right singular vectors of $\boldsymbol{M}$ corresponding to the sorted singular values.
- $\boldsymbol{U}_k \in \mathbb{R}^{m \times k}$ are the top $k \leq r$ left singular vectors of $\boldsymbol{M}$.

An application of SVD

- An important application of SVD in machine learning is dimensionality reduction.
- Principal Component Analysis (PCA) is one such dimensionality reduction technique.

Statement of PCA

1. Let us assume we are given a mean-centred data matrix $\boldsymbol{X} \in \mathbb{R}^{N \times m}$, that is $\sum_{i=1}^{m} \boldsymbol{x}_i = 0$.
2. Let $k \in [1, N]$ be a given parameter.
3. Define $\mathcal{P}_k$ as the set of N -dimensional rank-k orthogonal projection matrices.
4. PCA consists of projecting the N -dimensional input data onto the k-dimensional linear subspace that minimizes reconstruction error.
5. The reconstruction error is the sum of the squared L_2-distances between the original data and the projected data.
6. The PCA algorithm is completely defined by the orthogonal projection matrix solution $\boldsymbol{P}^*$ of the following minimization problem:

$$\min_{\boldsymbol{P} \in \mathcal{P}_k} ||\boldsymbol{P}\boldsymbol{X} - \boldsymbol{X}||_F^2$$

PCA theorem

Theorem (PCA)

Let $\boldsymbol{P}^* \in \mathcal{P}_k$ be the PCA solution, i.e. the orthogonal projection matrix solution of the minimization problem:

$$\min_{\boldsymbol{P} \in \mathcal{P}_k} ||\boldsymbol{PX} - \boldsymbol{X}||_F^2$$

Then $\boldsymbol{P}^* = \boldsymbol{U}_k \boldsymbol{U}_k^t$, where $\boldsymbol{U}_k \in \mathbb{R}^{N \times k}$ is the matrix formed by the top k singular vectors of $\boldsymbol{C} = \frac{1}{m}\boldsymbol{XX}^t$, the sample covariance matrix of coresponding to $\boldsymbol{X}$. Moreover, the associated k-dimensional representation of $\boldsymbol{X}$ is given by $\boldsymbol{Y} = \boldsymbol{U}_k^t \boldsymbol{X}$.

Illustrative Example

Example

In an experiment, four features were measured from a random sample of 10,000 human subjects. The sample covariance matrix was computed and the four eigenvalues were found to be 16.5, 5.4, 1.5 and 0.4. The eigenvectors corresponding to the first two eigenvalues were

$$\mathbf{u}_1^T = \begin{bmatrix} 0.39 & 0.42 & 0.44 & 0.69 \end{bmatrix}$$
$$\mathbf{u}_2^T = \begin{bmatrix} 0.40 & 0.39 & 0.42 & -0.72 \end{bmatrix}$$

1. What is the percentage of the variance in the original data explained by the first two principal components?
2. Assume that we decided to reduce the feature set to two (2). We need to transform (project) any new incoming 4-vector feature into a 2-vector corresponding to the two principal vectors we discovered in our eigen-analysis. The formula for this transformation is

$$\boldsymbol{x}_{\text{new}} = \boldsymbol{U}\boldsymbol{x}_{\text{old}}$$

   where

$$\boldsymbol{U} = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{bmatrix}$$

   What is the transformed vector corresponding to an input feature

$$\boldsymbol{x}^t = \begin{bmatrix} 0.5 & 0.6 & 1.6 & 0.9 \end{bmatrix}?$$

Solution I

1. The four given eigenvalues (16.5, 5.4, 1.5 and 0.4.) provide the amount of variance in the data explained by each of the principal components. Hence this is a straightforward question. We simply add the first two values and divide by the sum of the four values.

$$\begin{aligned}\text{Percent explained} &= \frac{16.5 + 5.4}{16.5 + 5.4 + 1.5 + 0.4} \times 100\% \\ &= \frac{21.9}{23.8} \times 100\% \\ &= 92.01\%\end{aligned}$$

2. The transformation matrix is given by the the eigenvectors corresponding to the largest two eigenvalues. We were given the vectors in the question;

$$\begin{aligned}\mathbf{u}_1^T &= \begin{bmatrix}0.39 & 0.42 & 0.44 & 0.69\end{bmatrix} \\ \mathbf{u}_2^T &= \begin{bmatrix}0.40 & 0.39 & 0.42 & -0.72\end{bmatrix}\end{aligned}$$

Hence our matrix $\boldsymbol{U}$ can be written as

$$\boldsymbol{U} = \begin{bmatrix}0.39 & 0.42 & 0.44 & 0.69 \\ 0.40 & 0.39 & 0.42 & -0.72\end{bmatrix}$$

Solution II

The transformation of the given 4-vector into a 2-vector is accomplished by

$$\boldsymbol{x}_{\text{new}} = \begin{bmatrix} 0.39 & 0.42 & 0.44 & 0.69 \\ 0.40 & 0.39 & 0.42 & -0.72 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.6 \\ 1.6 \\ 0.9 \end{bmatrix}$$

$$= \begin{bmatrix} 1.772 \\ 1.0412 \end{bmatrix}$$

The new 2-vector is the best projection of the given 4-vector into a 2-dimensional subspace. We can say the we have learned a 2-subspace that best represents our 4-vector space of features.

There are more linear algebraic results and they will be introduced as we need them!