Data Science

Matrices and Vectors

PILLAI COLLEGE OF ENGINEERING
Autonomous

Satishkumar L. Varma

Professor, Department of Information Technology
PCE, New Panvel
Scopus | Web of Science | Google Scholar | Google Site | Website

Matrices and Vectors: Outline

- Matrices and Vectors
- Eigenvalues & Eigenvectors
- Orthogonality
- Vector Norms

"Well planned milestone is a key to success."

Matrices and Vectors: Learning Objectives & Outcomes

- **Learning Objectives:** Course Instructor or Faculty aims
  - To explain Matrices and Vectors
  - Eigenvalues & Eigenvectors
  - Orthogonality
  - Vector Norms

Matrices and Vectors: Introduction

- An m×n (read "m by n") matrix, denoted by A,
- It is a rectangular array of entries or elements (numbers, or symbols representing numbers) enclosed typically by square brackets,
- where m is the number of rows and n the number of columns.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Matrices and Vectors: Introduction

- **A** is ***square*** if $m = n$.
- **A** is ***diagonal*** if all off-diagonal elements are 0, and not all diagonal elements are 0.
- **A** is the ***identity matrix*** (**I**) if it is diagonal and all diagonal elements are 1.
- **A** is the ***zero*** or ***null matrix*** (**0**) if all its elements are 0.
- The ***trace*** of **A** equals the sum of the elements along its main diagonal.
- Two matrices **A** and **B** are ***equal*** iff the have the same number of rows and columns, and $a_{ij} = b_{ij}$.

Matrices and Vectors: Introduction

- **A** is ***square*** if $m= n$.
- **A** is ***diagonal*** if all off-diagonal elements are 0, and not all diagonal elements are 0.
- **A** is the ***identity matrix*** (**I**) if it is diagonal and all diagonal elements are 1.
- **A** is the ***zero*** or ***null matrix*** (**0**) if all its elements are 0.
- The ***trace*** of **A** equals the sum of the elements along its main diagonal.
- Two matrices **A** and **B** are ***equal*** iff the have the same number of rows and columns, and $a_{ij} = b_{ij}$.
- The ***transpose*** $\mathbf{A}^T$ of an $m\times n$ matrix **A** is an $n\times m$ matrix obtained by interchanging the rows and columns of **A**.
- A square matrix for which $\mathbf{A}^T=\mathbf{A}$ is said to be ***symmetric***.
- Any matrix X for which **XA**=**I** and **AX**=**I** is called the ***inverse*** of **A**.
- Let c be a real or complex number (called a ***scalar***).
- The ***scalar multiple*** of c and matrix **A**, denoted $c\mathbf{A}$, is obtained by multiplying every elements of **A** by c. If $c = -1$, the scalar multiple is called the ***negative*** of **A**.

Matrices and Vectors: Introduction

- A ***column vector*** is an $m \times 1$ matrix:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

- A ***row vector*** is a $1 \times n$ matrix: $\mathbf{b} = [b_1, b_2, \cdots b_n]$
- A column vector can be expressed as a row vector by using the transpose:

$$\mathbf{a}^T = [a_1, a_2, \cdots, a_m]$$

Matrices and Vectors: Matrix Operations

- Matrices and Vectors:The ***sum*** of two matrices **A** and **B** (of equal dimension), denoted **A** + **B**, is the matrix with elements $a_{ij} + b_{ij}$.
- The ***difference*** of two matrices, **A**− **B**, has elements $a_{ij} - b_{ij}$.
- The ***product***, **AB**, of $m\times n$ matrix **A** and $p\times q$ matrix **B**, is an $m\times q$ matrix **C** whose (i,j)-th element is formed by multiplying the entries across the ith row of **A** times the entries down the jth column of **B**; that is,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{pj}$$

- The ***inner product*** (also called ***dot product***) of two vectors

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

- is defined as (Note that the inner product is a scalar.)

$$\begin{aligned} \mathbf{a}^T\mathbf{b} = \mathbf{b}^T\mathbf{a} &= a_1b_1 + a_2b_2 + \cdots + a_mb_m \\ &= \sum_{i=1}^{m} a_ib_i. \end{aligned}$$

Matrices and Vectors: Vectors and Vector Spaces

- A ***vector space*** is defined as a nonempty set V of entities called *vectors* and associated scalars that satisfy the conditions outlined in A through C below.
- A vector space is *real* if the scalars are real numbers; it is *complex* if the scalars are complex numbers.
- Condition A: There is in V an operation called *vector addition*, denoted $\mathbf{x} + \mathbf{y}$, that satisfies:
  - 1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for all vectors $\mathbf{x}$ and $\mathbf{y}$ in the space.
  - 2. $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$ for all $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$.
  - 3. There exists in V a unique vector, called the *zero vector*, and denoted $\mathbf{0}$, such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ and $\mathbf{0} + \mathbf{x} = \mathbf{x}$ for all vectors $\mathbf{x}$.
  - 4. For each vector $\mathbf{x}$ in V, there is a unique vector in V, called the *negation* of $\mathbf{x}$, and denoted $-\mathbf{x}$, such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$ and $(-\mathbf{x}) + \mathbf{x} = \mathbf{0}$.
- Condition B: There is in V an operation called *multiplication by a scalar* that associates with each scalar c and each vector $\mathbf{x}$ in V a unique vector called the *product* of c and $\mathbf{x}$, denoted by $c\mathbf{x}$ and $\mathbf{x}c$, and which satisfies:
  - 1. $c(d\mathbf{x}) = (cd)\mathbf{x}$ for all scalars c and d, and all vectors $\mathbf{x}$.
  - 2. $(c + d)\mathbf{x} = \mathrm{c}\mathbf{x} + \mathrm{d}\mathbf{x}$ for all scalars c and d, and all vectors $\mathbf{x}$.
  - 3. $\mathrm{c}(\mathbf{x} + \mathbf{y}) = \mathrm{c}\mathbf{x} + \mathrm{c}\mathbf{y}$ for all scalars c and all vectors $\mathbf{x}$ and $\mathbf{y}$.
- Condition C: $1\mathbf{x} = \mathbf{x}$ for all vectors $\mathbf{x}$.

Matrices and Vectors: Vectors and Vector Spaces

- We are interested particularly in real vector spaces of real $m\times1$ column matrices.
- We denote such spaces by $\Re^m$, with vector addition and multiplication by scalars being as defined
  - earlier for matrices. Vectors (column matrices) in $\Re^m$ are written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

- **Example**
- The vector space with which we are most familiar is the two-dimensional real vector space $\Re^2$, in which we make frequent use of graphical representations for operations such as vector addition, subtraction, and multiplication by a scalar. For instance, consider the two vectors

$$\mathbf{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

- Using the rules of matrix addition and subtraction we have

$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad \mathbf{a} - \mathbf{b} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

- The following figure shows the familiar graphical representation of the preceding vector operations, as well as multiplication of vector **a** by scalar $c = -0.5$.

Matrices and Vectors: Vectors and Vector Spaces

- Consider two real vector spaces V_0 and V such that:
- Each element of V_0 is also an element of V (i.e., V_0 is a *subset* of V).
- Operations on elements of V_0 are the same as on elements of V.
- Under these conditions, V_0 is said to be a ***subspace*** of V.
- A ***linear combination*** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an expression of the form; where the α's are scalars.

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n$$

- A vector **v** is said to be ***linearly dependent*** on a set, S, of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ if and only if **v** can be written as a linear combination of these vectors. Otherwise, **v** is ***linearly independent*** of the set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.
- A set S of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in V is said to ***span*** some subspace V_0 of V if and only if S is a subset of V_0 and every vector $\mathbf{v}_0$ in V_0 is linearly dependent on the vectors in S.
- The set S is said to be a ***spanning set*** for V_0.
- A ***basis*** for a vector space V is a linearly independent spanning set for V.
- The number of vectors in the basis for a vector space is called the ***dimension*** of the vector space.
- If, for example, the number of vectors in the basis is n, we say that the vector space is n-dimensional.

Matrices and Vectors: Vectors and Vector Spaces

- An important aspect of the concepts just discussed lies in the representation of any vector in $\Re^m$ as a ***linear combination*** of the basis vectors. For example, any vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

- in $\Re^3$ can be represented as a linear combination of the basis vectors

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \text{ and } \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Matrices and Vectors: Vector Norms

- A ***vector norm*** on a vector space V is a function that assigns to each vector **v** in V a nonnegative real number, called the ***norm*** of **v**, denoted by $\|\mathbf{v}\|$.
- By definition, the norm satisfies the following conditions:

$$
\begin{aligned}
&(1)\ \|\mathbf{v}\| > 0 \text{ for } \mathbf{v} \neq \mathbf{0};\ \|\mathbf{0}\| = 0,\\
&(2)\ \|c\mathbf{v}\| = |c|\|\mathbf{v}\| \text{ for all scalars } c \text{ and vectors } \mathbf{v}, \text{ and}\\
&(3)\ \|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.
\end{aligned}
$$

- There are numerous norms that are used in practice.
- In our work, the norm most often used is the so-called **2-*norm***, which, for a vector **x** in real $\Re^m$, space is defined as

$$\|\mathbf{x}\| = \left[x_1^2 + x_2^2 + \cdots + x_m^2\right]^{1/2}$$

- which is recognized as the *Euclidean distance* from the origin to point **x**;
  - this gives the expression the familiar name Euclidean norm.
- The expression also is recognized as the length of a vector **x**, with origin at point **0**.
- From earlier discussions, the norm also can be written as

$$\|\mathbf{x}\| = \left[\mathbf{x}^T\mathbf{x}\right]^{1/2}$$

Matrices and Vectors: Vector Norms

- The ***Cauchy-Schwartz*** inequality states that

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

- Another well-known result used in the book is the expression

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

- where θ is the angle between vectors **x** and **y**.
  - From these expressions it follows that the inner product of two vectors can be written as

$$\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos\theta$$

- Thus, the inner product can be expressed as a function of the norms of the vectors and the angle between the vectors.
- From the preceding results, two vectors in $\Re^m$ are ***orthogonal*** if and only if their inner product is zero.
  - Two vectors are ***orthonormal*** if, in addition to being orthogonal, the length of each vector is 1.
- From the concepts just discussed, we see that an arbitrary vector **a** is turned into a vector $\mathbf{a}_n$ of unit length by performing the operation $\mathbf{a}_n = \mathbf{a}/\|\mathbf{a}\|$. Clearly, then, $\|\mathbf{a}_n\| = 1$.
- A ***set of vectors*** is said to be an ***orthogonal*** set if every two vectors in the set are orthogonal.
  - A ***set of vectors*** is ***orthonormal*** if every two vectors in the set are orthonormal.

Matrices and Vectors: Orthogonality

- Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be an orthogonal or orthonormal basis in the sense defined in the previous section.
- Then, an important result in vector analysis is that any vector v can be represented with respect to the orthogonal basis B as

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n$$

- (where the coefficients are given by)

$$\alpha_i = \frac{\mathbf{v}^T \mathbf{v}_i}{\mathbf{v}_i^T \mathbf{v}_i} = \frac{\mathbf{v}^T \mathbf{v}_i}{\|\mathbf{v}_i\|^2}$$

- The key importance of this result is that,
  - if we represent a vector as a linear combination of orthogonal or orthonormal basis vectors,
    - we can determine the coefficients directly from simple inner product computations.
- It is possible to convert a linearly independent spanning set of vectors into an orthogonal spanning set by using the well-known ***Gram-Schmidt*** process.
- There are numerous programs available that implement the Gram-Schmidt and similar processes, so we will not dwell on the details here.

Matrices and Vectors: Eigenvalues & Eigenvectors

- **Definition:** The ***eigenvalues*** of a real matrix **M** are the real numbers λ for which there is a nonzero vector **e** such that

  $$\mathbf{M}\mathbf{e} = \lambda\, \mathbf{e}.$$

- The ***eigenvectors*** of **M** are the nonzero vectors **e** for which there is a real number λ such that $\mathbf{M}\mathbf{e} = \lambda\, \mathbf{e}$.
-
- If $\mathbf{M}\mathbf{e} = \lambda\, \mathbf{e}$ for $\mathbf{e} \neq 0$, then **e** is an ***eigenvector*** of **M** associated with ***eigenvalue*** λ, and vice versa. The eigenvectors and corresponding eigenvalues of **M** constitute the ***eigensystem*** of **M**.
-
- Numerous theoretical and truly practical results in the application of matrices and vectors stem from this beautifully simple definition.
- **Example:** Consider the matrix $\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

It is easy to verify that $\mathbf{M}\mathbf{e}_1 = \lambda_1 \mathbf{e}_1$ and $\mathbf{M}\mathbf{e}_2 = \lambda_2 \mathbf{e}_2$ for $\lambda_1 = 1$, $\lambda_2 = 2$ and

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- In other words, $\mathbf{e}_1$ is an eigenvector of **M** with associated eigenvalue λ_1, and similarly for $\mathbf{e}_2$ and λ_2.

Matrices and Vectors: Eigenvalues & Eigenvectors

- Essential Properties: of vectors and matrices. Assume a real matrix of order $m \times m$; applicable to complex numbers.
- **1.** If $\{\lambda_1, \lambda_2, \ldots, \lambda_q\}$, $q \leq m$, is set of distinct eigenvalues of **M**, and $\mathbf{e}_i$ is an eigenvector of **M** with corresponding eigenvalue λ_i, $i = 1,2,\ldots,q$, then $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_q\}$ is a linearly independent set of vectors. An important implication of this property: If an $m \times m$ matrix **M** has m distinct eigenvalues, its eigenvectors will constitute an orthogonal (orthonormal) set, which means that any m-dimensional vector can be expressed as a linear combination of the eigenvectors of **M**.
- **2.** The numbers along the main diagonal of a diagonal matrix are equal to its eigenvalues. It is not difficult to show using the definition $\mathbf{Me} = \lambda\,\mathbf{e}$ that the eigenvectors can be written by inspection when **M** is diagonal.
- **3.** A real, symmetric $m \times m$ matrix **M** has a set of m linearly independent eigenvectors that may be chosen to form an orthonormal set. This property is of particular importance when dealing with covariance matrices (e.g., see Section 11.4 and our review of probability) which are real and symmetric.
- **4.** A corollary of Property 3 is that the eigenvalues of an $m \times m$ real symmetric matrix are real, and the associated eigenvectors may be chosen to form an orthonormal set of m vectors.
- **5.** Suppose that **M** is a real, symmetric $m \times m$ matrix, and that we form a matrix **A** whose rows are the m orthonormal eigenvectors of **M**. Then, the product $\mathbf{AA}^T = \mathbf{I}$ because the rows of **A** are orthonormal vectors. Thus, we see that $\mathbf{A}^{-1} = \mathbf{A}^T$ when matrix **A** is formed in the manner just described.
- **6.** Consider matrices **M** and **A** in 5. The product $\mathbf{D} = \mathbf{AMA}^{-1} = \mathbf{AMA}^T$ is a diagonal matrix whose elements along the main diagonal are the eigenvalues of **M**. The eigenvectors of **D** are the same as the eigenvectors of **M**.

Matrices and Vectors: Eigenvalues & Eigenvectors

- **Example**
- Suppose that we have a random population of vectors, denoted by $\{\mathbf{x}\}$, with covariance matrix (see the review of probability):

$$\mathbf{C_x} = E\{(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})^T\}$$

- Suppose that we perform a transformation of the form $\mathbf{y} = \mathbf{Ax}$ on each vector $\mathbf{x}$, where the rows of $\mathbf{A}$ are the orthonormal eigenvectors of $\mathbf{C_x}$. The covariance matrix of the population $\{\mathbf{y}\}$ is

$$\begin{aligned}
\mathbf{C_y} &= E\{(\mathbf{y} - \mathbf{m_y})(\mathbf{y} - \mathbf{m_y})^T\} \\
&= E\{(\mathbf{Ax} - \mathbf{Am_x})(\mathbf{Ax} - \mathbf{Am_x})^T\} \\
&= E\{\mathbf{A}(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})^T\mathbf{A}^T\} \\
&= \mathbf{A}E\{(\mathbf{x} - \mathbf{m_x})(\mathbf{x} - \mathbf{m_x})^T\}\mathbf{A}^T \\
&= \mathbf{AC_xA}^T
\end{aligned}$$

- From Property 6, we know that $\mathbf{C_y}=\mathbf{AC_xA}^T$ is a diagonal matrix with the eigenvalues of $\mathbf{C_x}$ along its main diagonal.
  - The elements along the main diagonal of a covariance matrix are the variances of the components of the vectors in the population.
  - The off diagonal elements are the covariances of the components of these vectors.

Matrices and Vectors: Eigenvalues & Eigenvectors

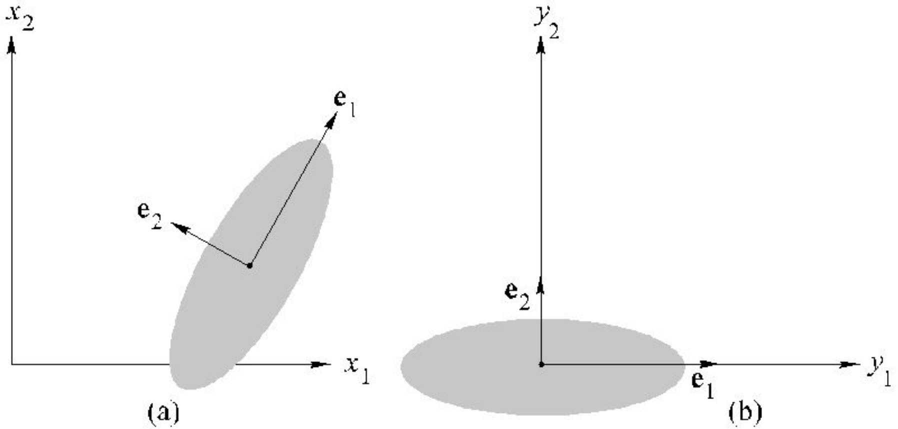
- **Example**
- The fact that $\mathbf{C}_\mathbf{y}$ is diagonal means that the elements of the vectors in the population $\{\mathbf{y}\}$ are ***uncorrelated*** (their covariances are 0).
  - Thus, we see that application of the linear transformation $\mathbf{y} = \mathbf{Ax}$ involving the eigenvectors of $\mathbf{C}_\mathbf{x}$ decorrelates the data, and the elements of $\mathbf{C}_\mathbf{y}$ along its main diagonal give the variances of the components of the **y**'s along the eigenvectors.
  - Basically, what has been accomplished here is a coordinate transformation that aligns the data along the eigenvectors of the covariance matrix of the population.
- The preceding concepts are illustrated in the following figure. Part (a) shows a data population $\{\mathbf{x}\}$ in two dimensions, along with the eigenvectors of $\mathbf{C}_\mathbf{x}$ (the black dot is the mean).
  - The result of performing the transformation $\mathbf{y}=\mathbf{A}(\mathbf{x} - \mathbf{m}_\mathbf{x})$ on the **x**'s is shown in Part (b) of the figure.

- **Example**

- The fact that we subtracted the mean from the **x**'s caused the **y**'s to have zero mean, so the population is centered on the coordinate system of the transformed data.

- It is important to note that all we have done here is make the eigenvectors the new coordinate system (y_1, y_2). Because the covariance matrix of the **y**'s is diagonal, this in fact also decorrelated the data.

- The fact that the main data spread is along $\mathbf{e}_1$ is due to the fact that the rows of the transformation matrix **A** were chosen according the order of the eigenvalues, with the first row being the eigenvector corresponding to the largest eigenvalue.



Summary

- Matrices and Vectors
- Eigenvalues & Eigenvectors

References

A. Text Books:

1. Eigenvalues - Wolfram Mathematica 8 Documentation
2. Basic Linear Algebra Concepts - Wiley Online Library
   https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118656747.app1

B. References:

3. R. C. Gonzalez & R. E. Woods; Digital Image Processing, 2nd ed. 2001

Thank You.