Mathematics for Machine Learning

Prof. Jaewook Lee
Statistical Learning and Computational Finance Lab.
Department of Industrial Engineering
jaewook@snu.ac.kr
http://slcf.snu.ac.kr

Chapter 0. Vectors and Matrices

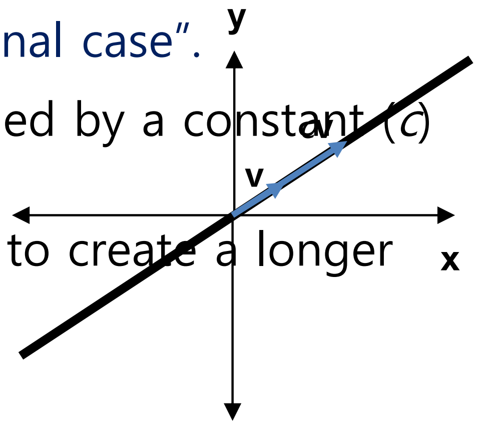
Prof. Jaewook Lee
Statistical Learning and Computational Finance Lab.
Department of Industrial Engineering
jaewook@snu.ac.kr
http://slcf.snu.ac.kr

Reading

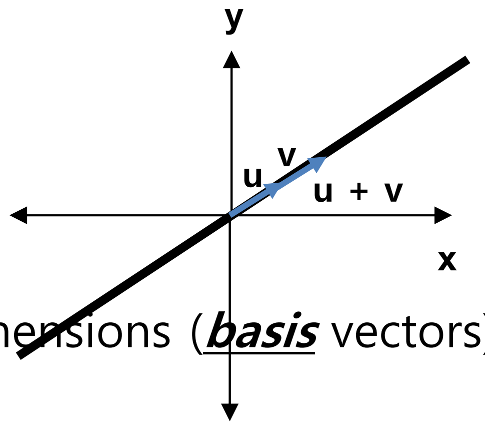
- [Strang. (2006), Chapter 1 Matrices and Gaussian Elimination]
- G. Strang *Linear Algebra And Its Applications-4th ed.* Cengage Learning, New York, 2006.
  - *Linear Algebra has become as basic and as applicable as calculus, and fortunately it is easier.* --Gilbert Strang, MIT
- Prof. Gilbert Strang's course videos:
  - http://ocw.mit.edu/OcwWeb/Mathematics/18-06Spring-2005/VideoLectures/index.htm
- Borrows some slides from S. Kalyanaraman, *Linear Algebra A gentle introduction*.

What is "Linear" & "Algebra"?

- Properties satisfied by a *line through the origin* ("one-dimensional case".
  - A directed arrow from the origin (**v**) on the line, when scaled by a constant (*c*) remains on the line
  - Two directed arrows (**u** and **v**) on the line can be "added" to create a longer directed arrow (**u** + **v**) in the same line.



- Wait a minute! This is nothing but *arithmetic with symbols*!
  - "Algebra": generalization and extension of arithmetic.
  - "Linear" operations: addition and scaling.



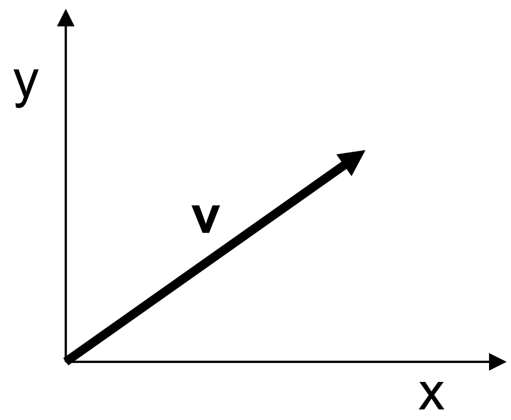
- Abstract and Generalize !
  - "Line" ↔ **vector space** having N dimensions
  - "Point" ↔ **vector** with N components in each of the N dimensions (***basis*** vectors).
    - Vectors have: "Length" and "Direction".
    - Basis vectors: "span" or define the space & its dimensionality.
  - Linear function transforming vectors ↔ **matrix**.
    - The function acts on each vector component and scales it
    - Add up the resulting scaled components to get a new vector!
    - In general: f(c**u** + d**v**) = cf(**u**) + df(**v**)

What is a Vector ?

- Think of a vector as a *directed line segment in N-dimensions*! (has “length” and “direction”)

- Basic idea: convert geometry in higher dimensions into algebra!
  - Once you define a “nice” *basis* along each dimension: x-, y-, z-axis …
  - Vector becomes a 1 x N matrix!
  - **v** = [a b c]^T
  - Geometry starts to become linear algebra on vectors like **v**!

$$\vec{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

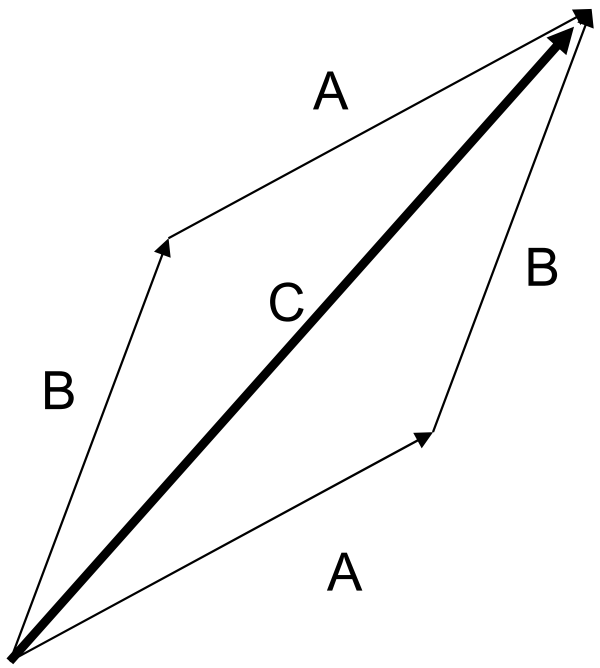


Examples of Geometry becoming Algebra

- Lines are vectors through the origin, scaled and translated: m**x** + **c**
  - Intersection of lines can be modeled as addition of vectors: solution of linear equations.
- Linear transformations of vectors can be associated with a matrix **A**, whose columns represent how each basis vector is transformed.
- Ellipses and conic sections:
  - $ax^2 + 2bxy + cy^2 = d$
  - Let $\mathbf{x} = [x\ y]^T$ and A is a symmetric matrix with rows $[a\ b]^T$ and $[b\ c]^T$
  - $\mathbf{x}^T\mathbf{A}\mathbf{x} = c$ {quadratic form equation for ellipse!}
    - This becomes convenient at higher dimensions
  - Note how a *symmetric* matrix A naturally arises from such a homogenous multivariate equation…

Vector Addition: A+B

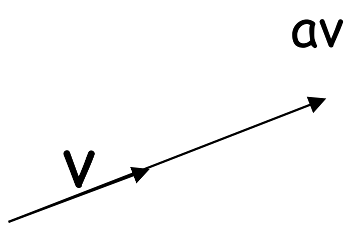
$$\mathbf{A+B} = (x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$$



A+B = C
(use the head-to-tail method to combine vectors)

Scalar Product: $a\mathbf{v}$

$$a\mathbf{v} = a(x_1, x_2) = (ax_1, ax_2)$$



Change only the length ("scaling"), but keep *direction fixed*.

Sneak peek: matrix operation (**Av**) can change *length*, *direction and also dimensionality*!

Vectors: Dot Product

$$A \cdot B = A^T B = \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} d \\ e \\ f \end{bmatrix} = ad + be + cf$$

Think of the dot product as a matrix multiplication

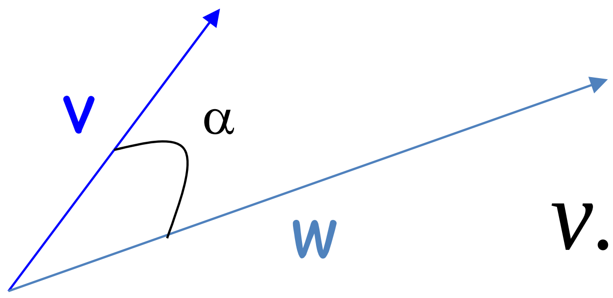
$$\|A\|^2 = A^T A = aa + bb + cc$$

The magnitude is the dot produ ct of a vector with itself

$$A \cdot B = \|A\| \, \|B\| \cos(\theta)$$

The dot product is also related to the a ngle between the two vectors

Inner (dot) Product: v.w or $w^T v$



$$v.w = (x_1, x_2).(y_1, y_2) = x_1 y_1 + x_2 . y_2$$

The inner product is a **SCALAR!**

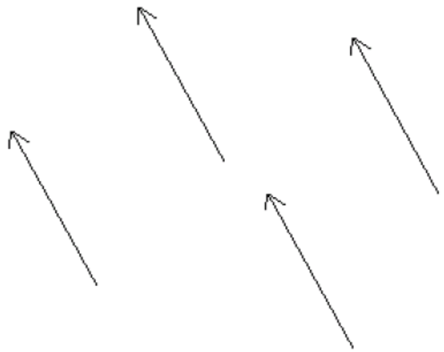
$$v.w = (x_1, x_2).(y_1, y_2) = \| v \| \cdot \| w \| \cos\alpha$$

$$v.w = 0 \Leftrightarrow v \perp w$$

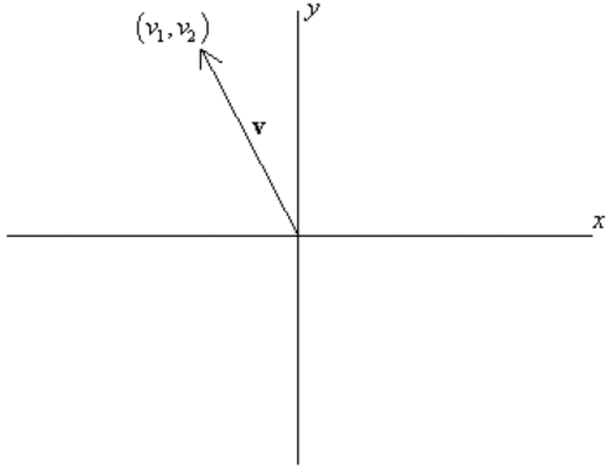
If vectors **v, w** are "columns", then dot product is $\mathbf{w}^T\mathbf{v}$

Bases & Orthonormal Bases

- Basis (or axes): frame of reference



VS



Basis: a space is totally defined by a set of vectors – any point is a *linear combination* of the basis

Ortho-Normal: orthogonal + normal

[**Sneak peek**:
Orthogonal: dot product is zero
Normal: magnitude is one]

$$x = [1 \quad 0 \quad 0]^T \qquad x \cdot y = 0$$
$$y = [0 \quad 1 \quad 0]^T \qquad x \cdot z = 0$$
$$z = [0 \quad 0 \quad 1]^T \qquad y \cdot z = 0$$

What is a Matrix?

- A matrix is a set of elements, organized into rows and columns

$$\begin{array}{cc} & \text{rows} \\ \text{columns} & \begin{bmatrix} a & b \\ c & d \end{bmatrix} \end{array}$$

Special matrices

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$ diagonal

$$\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$$ upper-triangular

$$\begin{pmatrix} a & b & 0 & 0 \\ c & d & e & 0 \\ 0 & f & g & h \\ 0 & 0 & i & j \end{pmatrix}$$ tri-diagonal

$$\begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix}$$ lower-triangular

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$ I (identity matrix)

Matrices

Matrix locations/size defined as rows x columns (R x C)

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}$$

d_{ij} : i^{th} row, j^{th} column

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

Square (3 x 3)

$$\begin{bmatrix} x11 & x12 & x13 \\ x21 & x22 & x23 \\ x31 & x32 & x33 \end{bmatrix}$$

3 dimensional (3 x 3 x 5)

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

Rectangular (3 x 2)

Basic Matrix Operations

- Addition, Subtraction, Multiplication: creating new matrices (or functio ns)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix}$$

Just add elements

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} - \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a-e & b-f \\ c-g & d-h \end{bmatrix}$$

Just subtract elements

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

Multiply each row b y each column

Example: Matrix Calculations

- Addition
  - Commutative: A+B=B+A
  - Associative: (A+B)+C=A+(B+C)

$$\mathbf{A}+\mathbf{B}=\begin{bmatrix}2 & 4\\ 2 & 5\end{bmatrix}+\begin{bmatrix}1 & 0\\ 3 & 1\end{bmatrix}=\begin{bmatrix}2+1 & 4+0\\ 2+3 & 5+1\end{bmatrix}=\begin{bmatrix}3 & 4\\ 5 & 6\end{bmatrix}$$

- Subtraction
  - By adding a negative matrix

$$\mathbf{A}-\mathbf{B}=\begin{bmatrix}2 & 4\\ 5 & 3\end{bmatrix}-\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}=\begin{bmatrix}2 & 4\\ 5 & 3\end{bmatrix}+\begin{bmatrix}-1 & -2\\ -3 & -4\end{bmatrix}=\begin{bmatrix}1 & 2\\ 2 & -1\end{bmatrix}$$

Basic Matrix Operations

- Transpose: You can think of it as
  - "flipping" the rows and columns

    OR
  - "reflecting" vector/matrix on line

e.g.

$$\begin{pmatrix} a \\ b \end{pmatrix}^T = \begin{pmatrix} a & b \end{pmatrix}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

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

Example: Transposition

$$\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \qquad \mathbf{b}^T = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} 3 & 4 & 9 \end{bmatrix} \qquad \mathbf{d}^T = \begin{bmatrix} 3 \\ 4 \\ 9 \end{bmatrix}$$

column $\longrightarrow$ row row $\longrightarrow$ column

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 4 & 1 \\ 6 & 7 & 4 \end{bmatrix} \qquad \mathbf{A}^T = \begin{bmatrix} 1 & 5 & 6 \\ 2 & 4 & 7 \\ 3 & 1 & 4 \end{bmatrix}$$

Matrix Multiplication

*"When **A** is a **m**x**n** matrix & **B** is a **k**x**l** matrix, **AB** is only possible if **n=k**. The result will be an **m**x**l** matrix"*

Simply put, can ONLY perform A*B IF:

Number of columns in A = Number of rows in B

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} A_1 & A_2 & A_3 \\ A_4 & A_5 & A_6 \\ A_7 & A_8 & A_9 \\ A_{10} & A_{11} & A_{12} \end{bmatrix} \end{array} \times \begin{array}{cc} l & \\ \begin{bmatrix} B_{13} & B_{14} \\ B_{15} & B_{16} \\ B_{17} & B_{18} \end{bmatrix} & k \end{array} = \mathbf{m} \times \mathbf{l} \text{ matrix}$$

Matrix Times Matrix

$$\mathbf{L} = \mathbf{M} \cdot \mathbf{N}$$

$$\begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \cdot \begin{bmatrix} n_{11} & n_{12} & n_{13} \\ n_{21} & n_{22} & n_{23} \\ n_{31} & n_{32} & n_{33} \end{bmatrix}$$

$$l_{12} = m_{11}n_{12} + m_{12}n_{22} + m_{13}n_{32}$$

Seoul National University

Multiplication

- Is AB = BA? Maybe, but maybe not!

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & \dots \\ \dots & \dots \end{bmatrix} \qquad \begin{bmatrix} e & f \\ g & h \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ea+fc & \dots \\ \dots & \dots \end{bmatrix}$$

- Matrix multiplication AB: apply transformation B first, and then again transfor m using A!
- Heads up: multiplication is NOT commutative!

- **Note**: If A and B both represent either pure "*rotation*" or "*scaling*" they can be interchanged (i.e. AB = BA)

Matrix multiplication

- Matrix multiplication is NOT commutative i.e the order matters!
  - $AB \neq BA$
- Matrix multiplication IS associative
  - $A(BC)=(AB)C$
- Matrix multiplication IS distributive
  - $A(B+C)=AB+AC$
  - $(A+B)C=AC+BC$

Identity matrix

- Identity matrix
  - A special matrix which plays a similar role as the number 1 in number multiplication?

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

For any nxn matrix $\boldsymbol{A}$, we have $\boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{I}_n\boldsymbol{A} = \boldsymbol{A}$
For any nxm matrix $\boldsymbol{A}$, we have $\boldsymbol{I}_n\boldsymbol{A} = \boldsymbol{A}$, and $\boldsymbol{A}\boldsymbol{I}_m = \boldsymbol{A}$ (so 2 possible matrices)

If the answers always A, why use an identity matrix?
Can't divide matrices, therefore to solve may problems have to use the inverse. The identity is important in these types of calculations.

Example: Identity matrix

Worked example
$\boldsymbol{A}\,\boldsymbol{I}_3 = \boldsymbol{A}$
for a 3x3 matrix:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1+0+0 & 0+2+0 & 0+0+3 \\ 4+0+0 & 0+5+0 & 0+0+6 \\ 7+0+0 & 0+8+0 & 0+0+9 \end{bmatrix}$$

Inverse of a matrix

- Inverse of a square matrix A, denoted by A^{-1} is the unique matrix s.t.
  - $AA^{-1} = A^{-1}A = I$ (identity matrix)
- If A^{-1} and B^{-1} exist, then
  - $(AB)^{-1} = B^{-1}A^{-1}$,
  - $\mathbf{(ABC)^{-1} = C^{-1}B^{-1}A^{-1}}$
  - $(A^T)^{-1} = (A^{-1})^T$
- For orthonormal matrices $\mathbf{A}^{-1} = \mathbf{A}^\mathsf{T}$
- For diagonal matrices

$$\mathbf{D}^{-1} = \mathrm{diag}\{d_1^{-1}, \ldots, d_n^{-1}\}$$

Dimensions

	Scalar	Vector	Matrix
Scalar	$\frac{\mathrm{d}y}{\mathrm{d}x}$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}x} = \left[\frac{\partial y_i}{\partial x}\right]$	$\frac{\mathrm{d}\mathbf{Y}}{\mathrm{d}x} = \left[\frac{\partial y_{ij}}{\partial x}\right]$
Vector	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y}{\partial x_j}\right]$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y_i}{\partial x_j}\right]$	
Matrix	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{X}} = \left[\frac{\partial y}{\partial x_{ji}}\right]$		

By Thomas Minka. Old and New Matrix Algebra Useful for Statistics

Examples

$$\frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T$$

$$\frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{x}$$

http://matrixcookbook.com/