Deep Learning

Linear Algebra

Kyungwoo Song

Relationship to Eigenvalues of the Hessian Li et al. (2018b) and Gur-Ari et al. (2019) argue that the first principal components of the SGD trajectory correspond to the top eigenvectors of the Hessian of the loss, and that these eigenvectors change slowly during training. This observation suggests that these principal components captures many of the sharp directions of the loss surface, corresponding to large Hessian eigenvalues. We expect then that our PCA subspace should include variation in the type of functions that it contains. See Appendix C for more details as well as a computation of Hessian and Fisher eigenvalues through a GPU accelerated Lanczos algorithm (Gardner et al., 2018).

2. Our Approach

Vanishing and exploding gradients can be controlled to a large extent by controlling the maximum and minimum *gain* of $\mathbf{W}$. The maximum gain of a matrix $\mathbf{W}$ is given by the spectral norm which is given by

$$||\mathbf{W}||_2 = \max \left[\frac{||\mathbf{W}\mathbf{x}||}{||\mathbf{x}||} \right]. \tag{6}$$

By keeping our weight matrix $\mathbf{W}$ close to orthogonal, one can ensure that it is close to a norm-preserving transformation (where the spectral norm is equal to one, but the minimum gain is also one). One way to achieve this is via a simple soft constraint or regularization term of the form:

$$\lambda \sum_i ||\mathbf{W}_i^T \mathbf{W}_i - \mathbf{I}||^2. \tag{7}$$

However, it is possible to formulate a more direct parameterization or factorization for $\mathbf{W}$ which permits *hard*

A Lie Group Generators

The six Lie group generating matrices for affine transformations in 2D are,

$$G_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad G_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$G_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad G_6 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Source: Learning Invariances in Neural Networks

Linear Algebra in ML

Overview

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} x_0 & x_{11} & x_{12} & \cdots & x_{1k} \\ x_0 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ x_0 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad \epsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

$n \times k$

- $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$
  - $\hat{Y} = X\hat{\beta}$
  - $\hat{\beta} = (X'X)^{-1}X'Y$
- 여기서 가장 어려운 부분이 무엇인가요?
  - $\hat{\beta} = (X'X)^{-1}X'Y$

if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

then $A^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

- 1) Invertible 할까요?
  - 즉, $X'X$의 역행렬이 존재할까요?
  - 우리는 역행렬이 존재하지 않는 행렬도 알고 있습니다.
  - 예시) $X = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ 라고 하면, X의 역행렬은 존재하지 않습니다.
  - 그렇다면, $X'X$는 어떨까요?
  - $X = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$이면, $X'X = \begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix}$이 되고, 그러면 $X'X$ 의 역행렬 또한 존재하지 않습니다.
  - **만약 X가 full-rank matrix이면, X^TX 는 역행렬이 존재합니다.**
    - ❖ 이런것들은 선형대수학 관련 내용을 다룰 때, 잠깐 같이 다룰 예정입니다.

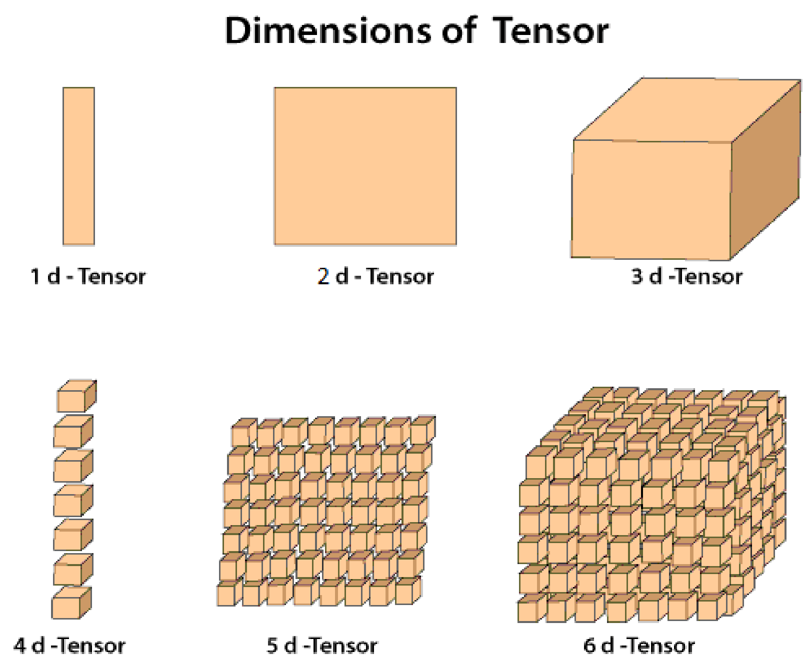
Contents

- Linear Algebra
  - Basic Operation
  - Vector Space
  - Orthogonal
  - Decomposition

Source:

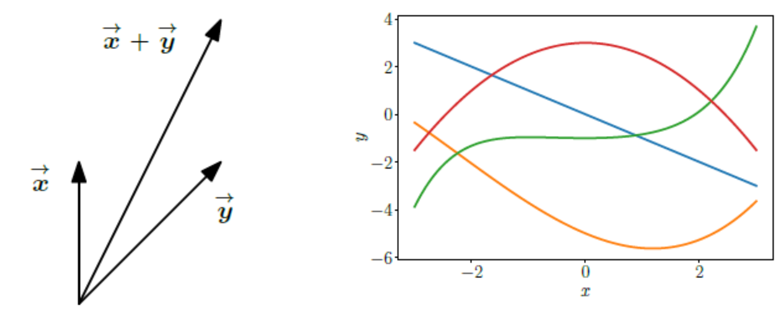
Linear Algebra – Basic Operation

- Scalar: single number
  - 0^{th} order tensor
  - Example) 1, -0.2, …
- Vector: array of numbers
  - 1^{st} order tensor
  - $x = \begin{bmatrix} 0.1 \\ 0.7 \\ -0.2 \end{bmatrix}$
- Matrix: 2-D array
  - 2^{nd} order tensor
  - $x = \begin{bmatrix} 0.1 & 9.0 \\ 0.7 & -1.4 \\ -0.2 & 2.1 \end{bmatrix}$
- Tensor
  - 만약 2차원보다 더 큰 차원을 표현하고 싶다면?



Source: https://stackoverflow.com/questions/64111559/how-to-understand-tensors-with-multiple-dimensions

- Linear Algebra (선형대수학)
  - Vector에 대한 내용을 다루는 학문
  - Vector를 변형하고 연산하는 것에 대한 규칙
- 여기서 Vector란?
  - 기하학적 의미: $\vec{x}$
  - 보다 일반적인 의미: 서로 더할 수 있고, scalar와 곱하여 새로운 것을 만들 수 있는 특별한 object
- 예시
  - Polynomial도 마찬가지로 Vector라고 볼 수 있음
    - ❖Polynomial끼리 더하여도, 여전히 polynomial임
    - ❖Polynomial와 scalar를 곱하여도, polynomial임
  - R^n의 원소도 vector라고 볼 수 있음
    - ❖$a = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in R^3$



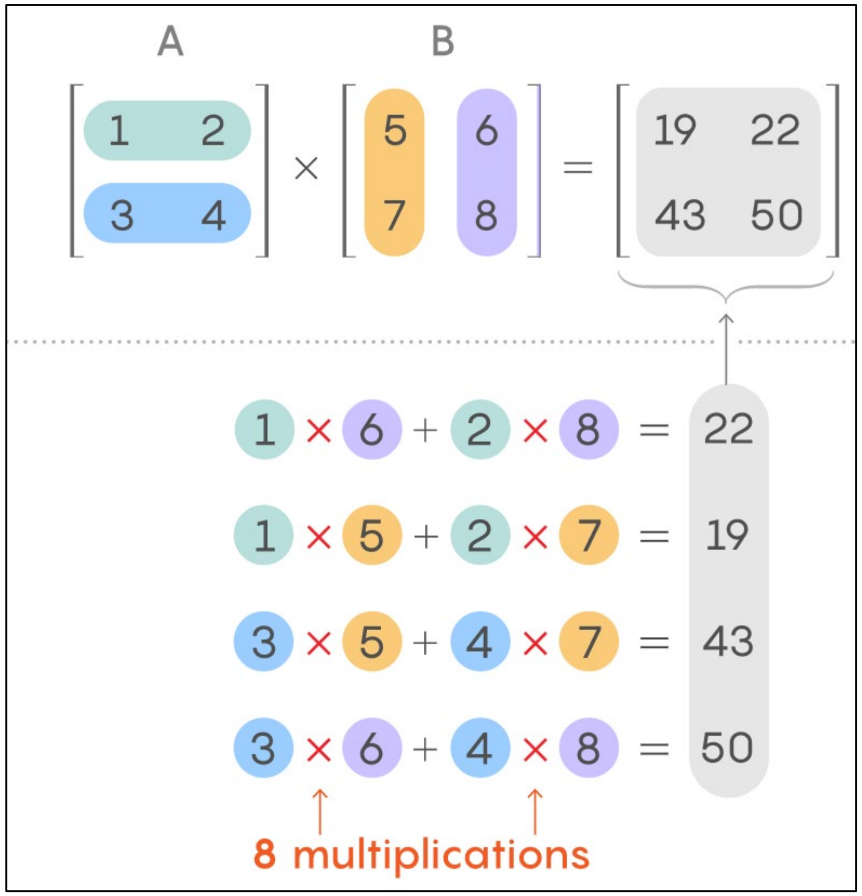
(a) Geometric vectors. (b) Polynomials.

Source: Mathematics for Machine Learning

Matrix Operation

Linear Algebra – Basic Operation

- Example
  - $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$
    - $2A - B = \begin{bmatrix} -3 & -2 \\ -1 & 0 \end{bmatrix}$, $AB = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$
- $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, identity matrix
  - $AI = IA = A$
- $\mathbf{0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, zero matrix
  - $A\mathbf{0} = \mathbf{0}A = \mathbf{0}$
- A^T: Transpose
  - $A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$
- $tr(A)$: trace
  - $tr(A) = 5$
- A^{-1}: Inverse Matrix
  - $AA^{-1} = A^{-1}A = I$ where $A^{-1} = \begin{bmatrix} -2 & -1 \\ 1.5 & -0.5 \end{bmatrix}$

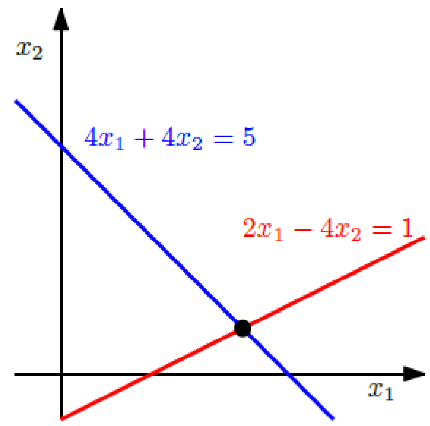


Source: https://www.quantamagazine.org/mathematicians-inch-closer-to-matrix-multiplication-goal-20210323/

- 아래와 같은 형태를 Systems of Linear Equations 라고 부름
  - $a_{11}x_1 + \cdots + a_{1n}x_n = b_1$

    $\dots$
  - $a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$
  - 위 식을 만족하는 $(x_1, \dots, x_n) \in R^n$: solution
  - $x_1, \dots, x_n$: unknown
- Example
  - $x_1 + x_2 + x_3 = 3$
  - $x_1 - x_2 + 2x_3 = 2$
  - $2x_1 \qquad + 3x_3 = 1$
  - 위 system of linear equations은 solution이 없음
    - 첫번째와 두번째 식을 더할 경우, 세번째 식과 상충됨

Source:



- Example
  - $x_1 + x_2 + x_3 = 3$
  - $x_1 - x_2 + 2x_3 = 2$
  - $x_2 + x_3 = 2$
  - 위 system of linear equations은 unique solution (1,1,1)을 가짐
- Example
  - $x_1 + x_2 + x_3 = 3$
  - $x_1 - x_2 + 2x_3 = 2$
  - $2x_1 \quad + 3x_3 = 5$
  - 위 system of linear equations은 다음과 같은 solution을 가짐
    - $\left(\frac{5}{2} - \frac{3}{2}a,\ \frac{1}{2} + \frac{1}{2}a,\ a\right), a \in R$
    - Infinitely many solutions
    - 1번식과 2번식을 더하면, 3번식과 동일함

Source:

- 아래와 같은 형태를 Systems of Linear Equations 라고 부름
  - $a_{11}x_1 + \cdots + a_{1n}x_n = b_1$

    $\ldots$
  - $a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$
  - 위 식을 만족하는 $(x_1, \ldots, x_n) \in R^n$: solution
  - $x_1, \ldots, x_n$: unknown
- 위를 보다 더 간결 하게 표현하면,
  - $x_1 \begin{bmatrix} a_{11} \\ \cdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ \cdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \cdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ \ldots \\ b_m \end{bmatrix}$
- 더욱 간결하게 표현하면, 아래와 같이 행렬이 나오게 됩니다.
  - $\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \ldots \\ b_m \end{bmatrix}$

Source:

Matrix Addition and Multiplication

Linear Algebra – Basic Operation

- 행렬은 본 교과목의 앞 부분에서 한번 살펴보았지만, 다시한번 복습 하겠습니다.
  - $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}, a_{ij} \in R$
  - (m, n) matrix
  - $(1, n)$ matrix: rows
  - $(m, 1)$ matrix: columns
- 행렬의 덧셈
  - $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \in R^{m \times n}, B = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \dots & \dots & \dots \\ b_{m1} & \dots & b_{mn} \end{bmatrix} \in R^{m \times n}$
  - $A + B = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \dots & \dots & \dots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix} \in R^{m \times n}$

Source:

- 행렬의 곱셈
  - $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \in R^{m\times n}$, $B = \begin{bmatrix} b_{11} & \dots & b_{1k} \\ \dots & \dots & \dots \\ b_{n1} & \dots & b_{nk} \end{bmatrix} \in R^{n\times k}$
  - $C = AB \in R^{m\times k}$
  - $c_{ij} = \sum_{l=1}^{n} a_{il}b_{lj}$ where $i = 1, \dots, m$ and $j = 1, \dots, k$
- Example
  - $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \in R^{2\times 3}$, $B = \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \in R^{3\times 2}$
  - $AB = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix}$ and $BA = \begin{bmatrix} 6 & 4 & 2 \\ -2 & 0 & 2 \\ 3 & 2 & 1 \end{bmatrix}$

NOTE
Hadamard Product (Element-wise product)
$A \odot B$

- Note
  - $c_{ij} \neq a_{ij}b_{ij}$
    - ❖$c_{ij} = a_{ij}b_{ij}$와 같이 연산하는 것을, Hadamard product 라고 부릅니다.

Source:

- Associativity
  - $(AB)C = A(BC)$
    - ❖ $A \in R^{m\times n}, B \in R^{n\times p}, C \in R^{p\times q}$
- Distributivity
  - $(A+B)C = AC + BC$
  - $A(C+D) \ = AC + AD$
    - ❖ $A, B \in R^{m\times n},\ C, D \in R^{n\times p}$

<u>Identity Matrix</u>

$$\boldsymbol{I}_n := \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

- Multiplication with the identity matrix
  - $I_m A = A I_n = A$

$$C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$
$$\lambda = 2$$
$$\Rightarrow \lambda C = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$$

Source:

Inverse Matrix

- $AB = BA = I$
  - The matrices A and B are inverse to each other
  - A 의 역행렬이 존재할 경우, A is called regular/invertible/nonsingular
  - A 의 역행렬이 존재하지 않는 경우, A is called singular/noninvertible

- Example
  - $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, B = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$
    - ❖ $a_{11}a_{22} - a_{12}a_{21} \neq 0$

- 여기서 $a_{11}a_{22} - a_{12}a_{21}$를 determinant 라고 부릅니다.
- det(A)

Source:

- 이러한 inverse matrix 는 몇 개나 있을까요?
- Theorem
  - If A is an invertible matrix, and if B and C are both inverse of A, then B=C; that is, an invertible matrix has a unique inverse.
- Proof)
  - B는 A의 역행렬이므로, BA=I
  - 양 변에 행렬 C를 곱하면, (BA)C=C
  - 즉, IC=C를 얻게 됨.
  - 만약 행렬 곱의 순서를 변경할 경우, B(AC)=C가 됨
  - C도 A의 역행렬이므로, BI=C, 즉 B=C가 됨

Source: Contemporary linear algebra

- For $A \in R^{m \times n}$, the matrix $B \in R^{n \times m}$ with $b_{ij} = a_{ji}$ is called the transpose of A
  - $B = A^T$

- A matrix A is symmetric if $A = A^T$

Note) Diagonal matrix
A matrix D is diagonal if and only if $D_{ij} = 0$ for all $i \neq j$

Inverting a square diagonal matrix is also efficient.

- Properties of inverse and transpose
  - $AA^{-1} = I = A^{-1}A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A + B)^{-1} \neq A^{-1} + B^{-1}$
  - $(A^T)^T = A$
  - $(A + B)^T = A^T + B^T$
  - $(AB)^T = B^T A^T$

Source: Contemporary linear algebra

- For $A \in R^{m \times n}$, the matrix $B \in R^{n \times m}$ with $b_{ij} = a_{ji}$ is called the transpose of A
  - $B = A^T$

- A matrix A is symmetric if $A = A^T$

- Theorem
  - If A is an invertible matrix, then A^T is also invertible and $(A^T)^{-1} = (A^{-1})^T$

- Proof) WTS: $(A^T)(A^{-1})^T = (A^{-1})^T A^T = I$
  - $A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$ (because $(AB)^T = B^T A^T$)
  - $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$ (because $(AB)^T = B^T A^T$)

Source: Contemporary linear algebra

- Definition
  - If A is a square matrix, then the **trace** of A, denoted by **tr(A)**, is defined to be sum of the entries on the main diagonal of A
- Example
  - $A = \begin{bmatrix} 3 & 2 \\ -1 & -8 \end{bmatrix}$
  - tr(A)=3+(-8) = -5

Note: anti-diagonal (“반”대각)

- Example
  - $u = \begin{bmatrix} u_1 \\ \cdots \\ u_n \end{bmatrix}$ and $v = \begin{bmatrix} v_1 \\ \cdots \\ v_n \end{bmatrix}$
  - $u^T v = [u_1 v_1 + \cdots + u_n v_n] = u \cdot v$
  - $uv^T = \begin{bmatrix} u_1 v_1 & \cdots & u_1 v_n \\ \cdots & \cdots & \cdots \\ u_n v_1 & \cdots & u_n v_n \end{bmatrix}$

$$u^T v = tr(uv^T)$$
$$u^T v = u \cdot v = v \cdot u = v^T u$$
$$tr(uv^T) = tr(vu^T) = u \cdot v$$

Source: Contemporary linear algebra

- Definition
  - If A is a square matrix, then the **trace** of A, denoted by **tr(A)**, is defined to be sum of the entries on the main diagonal of A

- Theorem) If A and B are square matrices with the same size, then:
  - $tr(A^T) = tr(A)$
  - $tr(cA) = c\ tr(A)$
  - $tr(A + B) = tr(A) + tr(B)$
  - $tr(A - B) = tr(A) - tr(B)$
  - $tr(AB) = tr(BA)$
    - ❖단, 일반적으로 $AB \neq BA$ 입니다!

- Quiz) 만약 A^{-1}가 존재할 때,
  - $tr(A)$와 $tr(A^{-1})$의 관계는 어떻게 될까요?

Source: Contemporary linear algebra

- Theorem
  - If r is a $1 \times n$ row vector and c is an $n \times 1$ column vector, then $rc = tr(cr)$
- Example
  - $r = [1,2]$ and $c = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$
  - $rc = 11$
  - $cr = \begin{bmatrix} 3 & 6 \\ 4 & 8 \end{bmatrix} \Rightarrow tr(cr) = 11$

$$\boxed{\begin{gathered} u^T v = tr(uv^T) \\ u^T v = u \cdot v = v \cdot u = v^T u \\ tr(uv^T) = tr(vu^T) = u \cdot v \end{gathered}}$$

- For $A \in R^{n \times n}$ and $u, v \in R^{n \times 1}$,
  - $Au \cdot v = v^T(Au) = (v^T A)u = (A^T v)^T u = u \cdot (A^T v)$
  - $u \cdot Av = (Av)^T u = (v^T A^T)u = v^T(A^T u) = A^T u \cdot v$
  - 즉, 행렬 A가 dot product 사이를 건너서 이동할 수 있다는 점입니다.

Source: Contemporary linear algebra

- For $A \in R^{n\times n}$ and $u, v \in R^{n\times 1}$,
  - $Au \cdot v = v^T(Au) = (v^T A)u = (A^T v)^T u = u \cdot (A^T v)$
  - $u \cdot Av = (Av)^T u = (v^T A^T)u = v^T(A^T u) = A^T u \cdot v$
  - 즉, 행렬 A가 dot product 사이를 건너서 이동할 수 있다는 점입니다.

- Example
  - $A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & 5 & 0 \\ -2 & 4 & 6 \end{bmatrix}, u = \begin{bmatrix} 3 \\ -2 \\ 4 \end{bmatrix}, v = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$
  - $Au \cdot v = u \cdot A^T v$

Note) Frobenius Norm

$\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$

$\|A\|_F = \sqrt{tr(AA^T)}$

Why?

$$\begin{aligned} \|A\|_F^2 &= \textstyle\sum_{i,j} A_{i,j}^2 \\ &= \textstyle\sum_i \left(\sum_j \ A_{ij}^T A_{ji}\right) \\ &= \textstyle\sum_i [A^T A]_{ii} \\ &= tr(A^T A) \end{aligned}$$

Source: Contemporary linear algebra

- System of Equations
  - $\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$
  - $x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 8 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} -4 \\ 12 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$
  - 2개의 방정식에, 4개의 변수가 존재
- Step1) Particular solution (또는 special solution)찾기
  - $(x_1, x_2, x_3, x_4) = (42,8,0,0)$ (단, 더 많은 solution이 있지 않을까?)
- Step2-1) 가능한 모든 solution 찾기
  - $x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 8 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} -4 \\ 12 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$ 찾기
  - $(x_1, x_2, x_3, x_4) = (8,2,-1,0)$
  - 여기에 λ_1을 곱해도 여전히 방정식은 만족됨

Source: Mathematics for Machine Learning

- System of Equations
  - $\begin{bmatrix}1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12\end{bmatrix}\begin{bmatrix}x_1 \\ x_2 \\ x_3 \\ x_4\end{bmatrix} = \begin{bmatrix}42 \\ 8\end{bmatrix}$
  - $x_1\begin{bmatrix}1 \\ 0\end{bmatrix} + x_2\begin{bmatrix}0 \\ 1\end{bmatrix} + x_3\begin{bmatrix}8 \\ 2\end{bmatrix} + x_4\begin{bmatrix}-4 \\ 12\end{bmatrix} = \begin{bmatrix}42 \\ 8\end{bmatrix}$
  - 2개의 방정식에, 4개의 변수가 존재
- Step2-2) 가능한 모든 solution 찾기
  - $x_1\begin{bmatrix}1 \\ 0\end{bmatrix} + x_2\begin{bmatrix}0 \\ 1\end{bmatrix} + x_3\begin{bmatrix}8 \\ 2\end{bmatrix} + x_4\begin{bmatrix}-4 \\ 12\end{bmatrix} = \begin{bmatrix}\mathbf{0} \\ \mathbf{0}\end{bmatrix}$ 찾기
  - $(x_1, x_2, x_3, x_4) = (-4, 12, 0, -1)$
  - 여기에 λ_2을 곱해도 여전히 방정식은 만족됨

Source: Mathematics for Machine Learning

- System of Equations
  - $\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$

비교적 풀기 쉬운 문제였습니다.
왜 그렇죠?
첫번째 column과 두번째 column
이 계산하기 편한 형태로 되어
있죠?

- Step3) 종합하기
  - $x_1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_3\begin{bmatrix} 8 \\ 2 \end{bmatrix} + x_4\begin{bmatrix} -4 \\ 12 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$
    - ❖ $(x_1, x_2, x_3, x_4) = (42,8,0,0)$
  - $x_1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_3\begin{bmatrix} 8 \\ 2 \end{bmatrix} + x_4\begin{bmatrix} -4 \\ 12 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$ 찾기
    - ❖ $(x_1, x_2, x_3, x_4) = (8,2,-1,0)$
    - ❖ 여기에 λ_1을 곱해도 여전히 방정식은 만족됨
  - $x_1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_3\begin{bmatrix} 8 \\ 2 \end{bmatrix} + x_4\begin{bmatrix} -4 \\ 12 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$ 찾기
    - ❖ $(x_1, x_2, x_3, x_4) = (-4,12,0,-1)$
    - ❖ 여기에 λ_2을 곱해도 여전히 방정식은 만족됨

Solution

$$\left\{ x = \begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + \lambda_1\begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} + \lambda_2\begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix}, \lambda_1, \lambda_2 \in R \right\}$$

Source: Mathematics for Machine Learning

- Elementary Transformations: Solution은 동일하게 유지해주면서, 보다 계산하기 편한 형태로 변환해주는 방식
  - 두 방정식을 교환 (두 row를 교환)
  - 특정 방정식 (row)에 상수 $\lambda \in R\backslash\{0\}$ 곱하기
  - 두 방정식 (row) 더하기
- Example
  - $-2x_1 + 4x_2 - 2x_3 - x_4 + 4x_5 = -3$
  - $4x_1 - 8x_2 + 3x_3 - 3x_4 + x_5 = 2$
  - $x_1 - 2x_2 + x_3 - x_4 + x_5 = 0$
  - $x_1 - 2x_2 \qquad - 3x_4 + 4x_5 = a$
  - $\Rightarrow \begin{bmatrix} -2 & 4 & -2 & -1 & 4 & | & -3 \\ 4 & -8 & 3 & -3 & 1 & | & 2 \\ 1 & -2 & 1 & -1 & 1 & | & 0 \\ 1 & -2 & 0 & -3 & 4 & | & a \end{bmatrix}$

Source: Mathematics for Machine Learning

- Elementary Transformations: Solution은 동일하게 유지해주면서, 보다 계산하기 편한 형태로 변환해주는 방식
  - 두 방정식을 교환 (두 row를 교환)
  - 특정 방정식 (row)에 상수 $\lambda \in R \backslash \{0\}$ 곱하기
  - 두 방정식 (row) 더하기
- Example
  - $\Rightarrow \left[\begin{array}{ccccc|c} -2 & +4 & -2 & -1 & +4 & -3 \\ +4 & -8 & +3 & -3 & +1 & +2 \\ +1 & -2 & +1 & -1 & +1 & +0 \\ +1 & -2 & +0 & -3 & +4 & +a \end{array}\right] \Rightarrow \left[\begin{array}{ccccc|c} +1 & -2 & +1 & -1 & +1 & +0 \\ +4 & -8 & +3 & -3 & +1 & +2 \\ -2 & +4 & -2 & -1 & +4 & -3 \\ +1 & -2 & +0 & -3 & +4 & +a \end{array}\right] \Rightarrow \left[\begin{array}{ccccc|c} +1 & -2 & +1 & -1 & +1 & +0 \\ +0 & +0 & -1 & +1 & -3 & +2 \\ +0 & +0 & +0 & -3 & +6 & -3 \\ +0 & +0 & -1 & -2 & +3 & +a \end{array}\right]$
  - $\Rightarrow \left[\begin{array}{ccccc|c} +1 & -2 & +1 & -1 & +1 & +0 \\ +0 & +0 & -1 & +1 & -3 & +2 \\ +0 & +0 & +0 & -3 & +6 & -3 \\ +0 & +0 & +0 & +0 & +0 & a+1 \end{array}\right] \Rightarrow \left[\begin{array}{ccccc|c} +1 & -2 & +1 & -1 & +1 & +0 \\ +0 & +0 & +1 & -1 & +3 & -2 \\ +0 & +0 & +0 & +1 & -2 & +1 \\ +0 & +0 & +0 & +0 & +0 & a+1 \end{array}\right]$

Swap
R_1 and R_3

두번째 row: $-4R_1$
세번째 row: $+2R_1$
네번째 row: $-R_1$

네번째 row:
$-R_2 - R_3$

두번째 row: $\times -1$
세번째 row: $\times -\frac{1}{3}$

$a = -1$

Source: Mathematics for Machine Learning

- Example
  - $$\begin{bmatrix} +1 & -2 & +1 & -1 & +1 & | & +0 \\ +0 & +0 & +1 & -1 & +3 & | & -2 \\ +0 & +0 & +0 & +1 & -2 & | & +1 \\ +0 & +0 & +0 & +0 & +0 & | & a+1 \end{bmatrix}$$
  - Particular solution (special solution)
    - $(x_1, x_2, x_3, x_4, x_5) = (2,0,-1,1,0)$
  - General solution
    - $$\left\{ x = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ 0 \\ -1 \\ 2 \\ 1 \end{bmatrix}, \lambda_1, \lambda_2 \in R \right\}$$

$$\underbrace{\mathrm{B} = \begin{pmatrix} \boxed{1} & 0 & 1 & 2 \\ 0 & \boxed{2} & 3 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{\text{REF}} \xrightarrow{\text{R2}=(\frac{1}{2})\text{R2}} \underbrace{\begin{pmatrix} \boxed{1} & 0 & 1 & 2 \\ 0 & \boxed{1} & \frac{3}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix}}_{\text{RREF}}$$

Note) REF and RREF

- **Row Echelon Form (REF)**
  - All nonzero rows are above any rows of all zeros
  - Each leading (nonzero) entry of a row is in a column to the right of the leading (nonzero) entry of the row above it
  - All entries in a column below a leading (nonzero) entry are zeros
- Reduced Row Echelon Form (RREF)
  - REF
  - The leading (nonzero) entry in each row is 1.
  - Each leading 1 is the only nonzero entry in its column.

Source: Mathematics for Machine Learning, https://www.math.fsu.edu/~cstover/teaching/sp18_mas3105/handouts/ch1/REFetc.pdf

Linear Algebra – Vector Space

쉽게 생각해서, 두 vector 의 합, 그리고 scalar 와의 곱이 잘 정의된 공간입니다.

- Definition
  - A **vector space** is a set V **along with an addition on V and a scalar multiplication on V** such that the following properties hold:
  - Commutativity
    - $u + v = v + u$ for all $u, v \in V$
  - Associativity
    - $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and scalar $a, b \in F$
  - Additive identity
    - There exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$
  - Additive inverse
    - For every $v \in V$, there exists $w \in V$ such that $v + w = 0$
  - Multiplicative identity
    - $1v = v$ for all $v \in V$
  - Distributive properties
    - $a(u + v) = au + av$ and $(a + b)v = av + bv$ for all $a, b \in F$ and all $u, v \in V$

F?!

Source: Linear Algebra Done Right

Example

- $\{0\}$ is a vector space
- $R^2 \backslash \{0\}$ is not a vector space

- $\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| x + y + z = 0 \right\}$ is a vector space if "+" and "·" are interpreted as
  - $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix}$ and $r \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} rx \\ ry \\ rz \end{pmatrix}$

- $\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| x + y + z = 1 \right\}$ is not a vector space if "+" and "·" are interpreted as
  - $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix}$ and $r \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} rx \\ ry \\ rz \end{pmatrix}$

Source: https://en.wikibooks.org/wiki/Linear_Algebra/Definition_and_Examples_of_Vector_Spaces

- 앞에서 살펴본 scalar multiplication은 결국 F에 depend 합니다.
  - 우리는 보다 정확하게 표현하기 위해서, V is a vector space over F 라고 많이 표현합니다.

- Definition
  - Elements of a vector space are called vectors or points

- Definition
  - A vector space over R is called a real vector space
  - A vector space over C is called a complex vector space

Source: Linear Algebra Done Right

Smallest subspace of V: $\{0\}$

- Definition
  - A subset U of V is called a **subspace** of V if U is also a vector space
    - (using the same addition and scalar multiplication as on V)

- 쉽게 생각해서, 부분집합인데, 잘 정의된 부분집합 입니다.
- 즉, U만으로도 vector space가 되어야 합니다.
  - Additive identity
    - $0 \in U$
  - Closed under addition
    - $u, w \in U$ implies $u + w \in U$
    - 즉, U는 V의 subset 인데, u, w 만 데려오고, $u + w$ 는 안데려오면, U는 vector space라고 할 수 없겠죠?
  - Closed under scalar multiplication
    - $a \in F$ and $u \in U$ implies $au \in U$

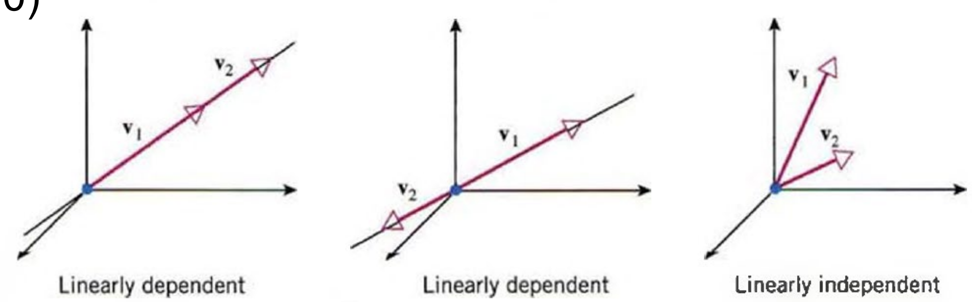
Source: Linear Algebra Done Right

F는 R 또는 C 입니다.

- Definition
  - A **linear combination** of a list $v_1, \ldots, v_m$ of vectors in V is a vector of the form $a_1 v_1 + \cdots + a_m v_m$ where $a_1, \ldots, a_m \in F$
- Example
  - $(17, -4,2)$ 는 $(2,1, -3)$ 과 $(1, -2,4)$의 linear combination
    - ❖ $(17, -4,2) = 6(2,1, -3) + 5(1, -2,4)$
  - $(17, -4,5)$ 는 $(2,1, -3)$ 과 $(1, -2,4)$의 linear combination 이 아니다.
    - ❖ $(17, -4,5) = a_1(2,1, -3) + a_2(1, -2,4)$를 만족하는 a_1과 a_2가 존재하지 않음
- Definition
  - The set of all linear combinations of a list of vectors $v_1, \ldots, v_m$ in V is called the **span** of $v_1, \ldots, v_m$, denoted span $(v_1, \ldots, v_m)$. In other words, $span(v_1, \ldots, v_m) = \{a_1 v_1 + \cdots + a_m v_m : a_1, \ldots, a_m \in F\}$.
- Example
  - $(17, -4,2) \in span((2,1, -3), (1, -2,4))$
  - $(17, -4,5) \notin span((2,1, -3), (1, -2,4))$

Source: Linear Algebra Done Right

- Definition
  - A list $v_1, \dots, v_m$ of vectors in V is called **linearly independent** if the only choice of $a_1, \dots, a_m \in F$ that makes $a_1 v_1 + \cdots + a_m v_m$ equal 0 is $a_1 = \cdots = a_m = 0$
- Definition
  - A list $v_1, \dots, v_m$ of vectors in V is called **linearly dependent** if there exist $a_1, \dots, a_m \in F$, not all 0, such that $a_1 v_1 + \cdots + a_m v_m = 0$
- Example
  - (1,0,0,0), (0,1,0,0), (0,0,1,0) is linearly independent in F^4
  - (2,3,1), (1,-1,2), (7,3,8) is linearly dependent in F^3
    - ❖2(2,3,1) + 3(1,-1,2) + (-1) (7,3,8) = (0,0,0)



Source: Linear Algebra Done Right, Contemporary linear algebra

- x_1, x_2, x_3 는 linearly independent 일까요 아닐까요?
  - $x_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}, x_3 = \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix}$
- 우리는 무엇을 확인 해봐야할까요? 정의를 기억해보세요!
  - $\lambda_1 \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix} + \lambda_3 \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix} = 0$ 를 만족하는 λ_i가 0뿐인지를 확인해야 합니다.
  - 앞에서 배웠던 Elementary Transformations을 활용해볼까요?
  - $\begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & -2 \\ -3 & 0 & 1 \\ 4 & 2 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ 즉, $\lambda_1 = \lambda_2 = \lambda_3 = 0$이 유일한 해 입니다.
  - Linearly independent

Source:

- Definition
  - A **basis** of V is a list of vectors in V that is linearly independent and spans V

- Example
  - The list $(1,0,\dots,0),(0,1,\dots,0),\dots,(0,\dots.,0,1)$ is a basis of F^n, called the standard basis of F^n
  - The list $(1,2),(3,5)$ is a basis of F^2
    - ❖ $\begin{bmatrix}1 & 3\\ 2 & 5\end{bmatrix} \Rightarrow \begin{bmatrix}1 & 3\\ 0 & -1\end{bmatrix} \Rightarrow \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} \Rightarrow \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$
  - The list $(1,2),(3,5),(4,13)$ spans F^2 but is not a basis of F^2
    - ❖ Not linearly independent

Source: Linear Algebra Done Right

- Example)
  - For a vector subspace $U \subset R^5$, spanned by the vectors x_1, x_2, x_3, x_4
    - ❖ $x_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, x_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \\ 2 \end{bmatrix}, x_3 = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, x_4 = \begin{bmatrix} -1 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix}$
    - ❖ 이 중, U의 basis는 무엇일까?
  - 우리가 check 해야 하는 것: $\sum \lambda_i x_i = 0$
    - ❖ $[x_1, x_2, x_3, x_4] = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & -4 & 8 \\ -1 & 1 & 3 & -5 \\ -1 & 2 & 5 & -6 \\ -1 & 2 & -3 & 1 \end{bmatrix} \Rightarrow \cdots \Rightarrow \begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
    - ❖ $\Rightarrow x_1, x_2, x_4$ are linearly independent
  - $\Rightarrow \{x_1, x_2, x_4\}$ is a basis of U

Source:

- Definition (**Rank**)
  - The number of linearly independent columns of a matrix $A \in R^{m \times n}$
  - The number of linearly independent rows of a matrix $A \in R^{m \times n}$
  - rank(A)

- Example
  - $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$
    - ❖ A 는 두개의 linearly independent rows 또는 columns를 가지고 있습니다.
    - ❖ Rank(A) = 2
  - $A = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$
    - ❖ Rank(A) = 2

Source:

- Properties of Rank
  - rank(A)=rank(A^T)
  - The rank of a full-rank matrix is the lesser of the number of rows and columns
    - ❖rank(A)=min(m, n)
    - ❖Full-rank matrix: rank equals the largest possible rank for a matrix
  - For all $A \in R^{n \times n}$, it holds that A is regular (invertible) if and only if rank(A)= n
- Proof
  - (⇒) A의 역행렬 B가 존재한다고 가정합니다. AB=BA=I
  - 그러면, AX=0을 만족하는 해를 찾아보겠습니다.
  - BAX=0이므로, X=0 입니다.
  - 즉, AX=0의 유일한 해는 X=0이므로, A는 full-rank 입니다.
  - (⇐) (?)

Source:

- Recall
  - $AB = BA = I$
    - ❖The matrices A and B are inverse to each other
    - ❖A 의 역행렬이 존재할 경우, A is called regular/invertible/nonsingular
    - ❖A 의 역행렬이 존재하지 경우, A is called singular/noninvertible
  - Example
    - ❖$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, B = \frac{1}{a_{11}a_{22}-a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$
      - ➢ $a_{11}a_{22} - a_{12}a_{21} \neq 0$

- 자 그러면, 더 큰 행렬의 matrix inverse는 어떻게 구할 수 있을까요?

Source:

- 앞에서 배웠던 Elementary Transformations 를 다시 보겠습니다.
- Elementary Transformations: Solution은 동일하게 유지해주면서, 보다 계산하기 편한 형태로 변환해주는 방식
  - 두 방정식을 교환 (두 row를 교환)
  - 특정 방정식 (row)에 상수 $\lambda \in R\backslash\{0\}$ 곱하기
  - 두 방정식 (row) 더하기
- 위 과정들은, 행렬로 표현 가능합니다.
  - $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$: 두번째 행과 세번째 행을 교환하기
  - $\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$: 두번째 행에 -3 곱하기
  - $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$: 세번째 행에 3을 곱해서, 첫번째 행에 더하기

Source: Contemporary linear algebra

- Example)
  - $A = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{bmatrix}$
  - $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}$
  - $\Rightarrow EA = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 5 & 4 & 12 & 12 \end{bmatrix}$

- 역행렬을 구하기 위한 기본적인 아이디어
  - $Ax = I$
  - 만약 역행렬이 존재한다면, $\boldsymbol{x} = \boldsymbol{A}^{-1}$
  - $E_1 E_2 \dots E_n A x = E_1 E_2 \dots E_n I$
  - $E_1 E_2 \dots E_n A$를 I로 만든다면, $\boldsymbol{x} = \boldsymbol{E_1 E_2 \dots E_n I}$

즉, elementary transformation 을 통해서, Identity를 만들면 된다!

Source:

Elementary Transformations

- $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}$의 inverse matrix 구하기

- $\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 0 & 8 & 0 & 0 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & -2 & 5 & -1 & 0 & 1 \end{array}\right]$

첫번째 행에 -2 곱하고,
두번째 행에 더해주고,
첫번째 행에 -1 곱하고,
세번째행에 더해주기

- $\Rightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & -1 & -5 & 2 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right]$

두번째행에 2 곱해서
세번째행에 더하기
+ 세번째행에 -1 곱셈

- $\Rightarrow \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & -14 & 6 & 3 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -40 & 16 & 9 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right]$

세번째 행에 3 곱하고 두번째 행에 덧셈
세번째 행에 -3 곱하고 첫번째 행에 덧셈

두번째행에 -2 곱해서
첫번째행에 더하기

Source:

Linear Algebra – Orthogonal

- Norm
  - Motivation
    - 어떻게 하면 주어진 벡터,행렬 등의 길이 또는 크기를 측정할 수 있을까?
  - Example
    - The length of a vector x in R^2 or R^3
    - $\|x\| = \sqrt{x_1^2 + x_2^2}$

겁먹지 마시고, Vector간의 덧셈연산과 Vector, scalar간의 곱셈 연산이 잘 정의된 좋은 공간이라고 생각합시다!

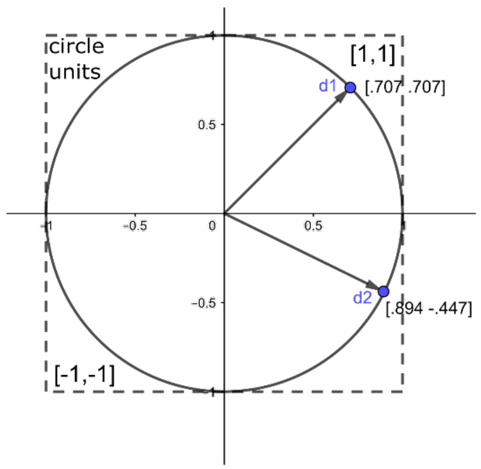
  - Definition
    - A norm on a vector space V is a function, $\|\cdot\|: V \to R,\ x \to \|x\|$ such that for all $\lambda \in R$ and $x, y \in V$ the following hold:
      - Absolutely homogenous: $\|\lambda x\| = |\lambda|\|x\|$
      - Triangle inequality: $\|x + y\| \le \|x\| + \|y\|$
      - Positive definite: $\|x\| \ge 0$ and $\|x\| = 0$ if and only if $x = 0$
  - Another Example
    - $\|x\|_p = (\sum_{i \in I} |x_i|^p)^{1/p}$ for $1 \le p \le \infty$
      - Note: $\|x\|_\infty = \max\limits_{1 \le i \le n} |x_i|$
    - $\|x\|_2$은 많이 보던 것이죠?! (Euclidean norm, L2 norm)
  - 별 다른 말이 없으면, vector norm 으로 생각하시면 됩니다.
    - Matrix norm은 추후에…
      - Note: $\|A\|_p = \sup\limits_{x \ne 0} \frac{\|Ax\|_p}{\|x\|_p}$

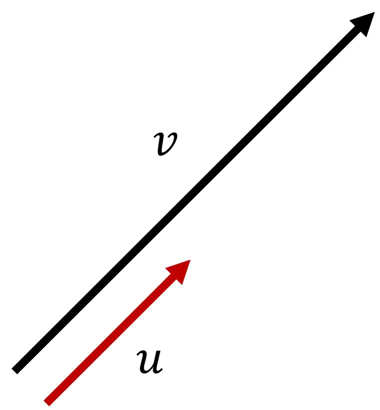
Note) Frobenius Norm

$$\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$$

Source: Linear Algebra Done Right

- Unit vector
  - Unit vector 는, vector 의 norm 이 1인 vector 를 의미합니다.
  - 다음과 같이, unit vector 를 만들 수 있습니다.
    - ❖ $u = \frac{1}{\|v\|}v$
  - 왜 그럴까요? $\|v\| = k$라고 가정하겠습니다.
    - ❖ $\|u\| = \left\|\frac{1}{\|v\|}v\right\| = \left\|\frac{1}{k}v\right\| = \frac{1}{k}\|v\| = \frac{1}{k}k = 1$





Source: Contemporary linear algebra https://en.wikipedia.org/wiki/Unit_vector

- 두 vector 간의 similarity 를 측정할 때, 우리는 dot product, 또는 Euclidean inner product 를 많이 활용합니다.

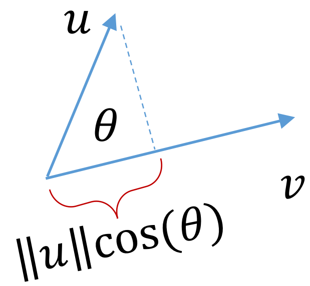
- Definition
  - If $u = (u_1, \dots, u_n)$ and $v = (v_1, \dots, v_n)$ are vectors in R^n, then the dot product of u and v, also called the Euclidean inner product of u and v, is denoted by $u \cdot v$ and is defined by the formula $u \cdot v = u_1 v_1 + \cdots + u_n v_n$

- 더 나아가, cosine similarity로 많이 측정합니다.
  - $\cos\theta = \frac{u \cdot v}{\|u\|\|v\|}$
  - 위의 dot product와 무엇이 다른지 감이 오시나요?
  - Normalization term 이 존재하는 것을 볼 수 있습니다.
  - 즉, unit vector 로 만든 다음에, dot product 를 진행하는것과 동일하죠?
  - Cosine similarity 의 max와 min 값은 어떻게 될까요?

Source: Contemporary linear algebra https://en.wikipedia.org/wiki/Unit_vector

Orthogonal

- 만약 두 vector 간의 dot product 로 similarity 를 측정하였는데 0이 나오면 어떤 의미일까요?
  - $u \cdot v = \|u\|\|v\| \cos\theta$ 입니다.
  - 즉, 0이라는 의미는, u 또는 v의 크기가 0이거나, $\cos\theta = 0$ 인 경우 입니다.
  - 만약 u와 v의 크기가 0이 아니라고 하면, θ가 90도 라는 의미입니다.
  - ⇒ orthogonal



- Definition
  - Two vectors u and v in R^n are said to be **orthogonal** if $u \cdot v = 0$, and a nonempty set of vectors in R^n is said to be an **orthogonal set** if each pair of distinct vectors in the set is orthogonal
- Example
  - $v_1 = (1,2,2,4)$, $v_2 = (-2,1,-4,2)$, $v_3 = (-4,2,2,-1)$ form an orthogonal set in R^4

Source: Contemporary linear algebra https://en.wikipedia.org/wiki/Unit_vector

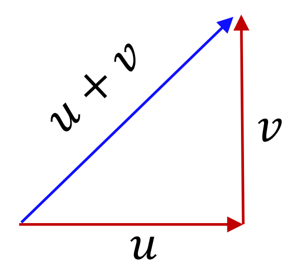
- Orthogonal 에서 조금만 더 나아가겠습니다. Orthogonal 인데, length (norm, magnitude)가 1인 경우를 orthonormal 이라고 부릅니다.

- Definition
  - Two vectors u and v in R^n are said to be **orthonormal** if they are orthogonal and have length 1, and a set of vectors is said to be an **orthonormal set** if every vector in the set has length 1 and each pair of distinct vectors is orthogonal

- Example
  - $q_1 = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}, \frac{4}{5}\right), q_2 = \left(-\frac{2}{5}, \frac{1}{5}, -\frac{4}{5}, \frac{2}{5}\right), q_3 = \left(-\frac{4}{5}, \frac{2}{5}, \frac{2}{5}, -\frac{1}{5}\right)$ form an orthonormal set in R^4

- Theorem (Theorem of Pythagoras)
  - If u and v are orthogonal vectors in R^n, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$

- Proof
  - $\|u + v\|^2 = (u + v) \cdot (u + v) = \|u\|^2 + 2(u \cdot v) + \|v\|^2 = \|u\|^2 + \|v\|^2$



Source: Contemporary linear algebra https://en.wikipedia.org/wiki/Unit_vector

- Theorem (Cauchy-Schwarz Inequality)
  - If u and v are vectors in R^n, then $(u \cdot v)^2 \leq \|u\|^2 \|v\|^2$ or equivalently, $|u \cdot v| \leq \|u\| \|v\|$
- Proof)
  - $z = \frac{u}{\|u\|} - \frac{v}{\|v\|} \Rightarrow z \cdot z = \frac{u \cdot u}{\|u\|^2} + \frac{v \cdot v}{\|v\|^2} - \frac{2u \cdot v}{\|u\|\|v\|} = 2 - \frac{2u \cdot v}{\|u\|\|v\|} \geq 0$
  - $\Rightarrow \|u\|\|v\| \geq u \cdot v$

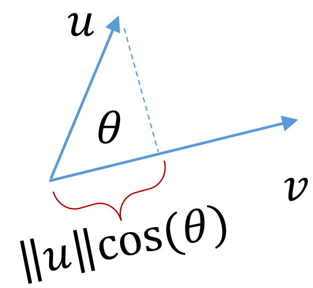
- Theorem (Triangle Inequality)
  - If u, v and w are vectors in R^n, then $\|u + v\| \leq \|u\| + \|v\|$
- Proof)
  - $\|u + v\|^2 = (u + v) \cdot (u + v) = \|u\|^2 + 2(u \cdot v) + \|v\|^2$
  - By the Cauchy-Schwarz inequality, $u \cdot v \leq \|u\|\|v\|$
  - $\Rightarrow \|u + v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$
  - $\Rightarrow \|u + v\| \leq \|u\| + \|v\|$

Source: Contemporary linear algebra https://en.wikipedia.org/wiki/Unit_vector

- Definition
  - A square matrix A is said to be **orthogonal** if $A^{-1} = A^T$
  - 즉, $A^TA = AA^T = I$ 라는 뜻입니다.
  - 즉, 행렬의 row 또는 column 이 orthonormal vectors 로 이뤄진 것을 의미합니다.
- Example
  - $A = \begin{bmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{bmatrix}$라고 하면, $AA^T = I$
- Property
  - Orthogonal Matrix의 transpose 도 orthogonal 입니다.
  - Orthogonal Matrix의 inverse matrix 도 orthogonal 입니다.
  - Orthogonal Matrix의 곱도, orthogonal matrix 입니다.
    - ❖ $(AB)^{-1} = B^{-1}A^{-1} = B^TA^T = (AB)^T$

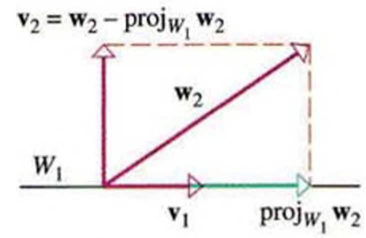
Source:

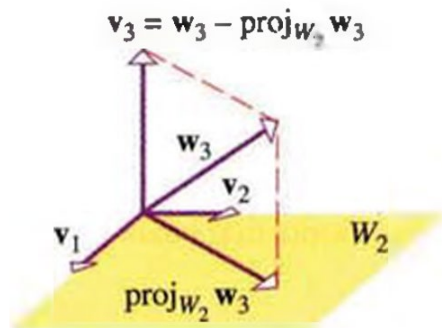
- Project u onto subspace V spanned by v
  - $proj_V u = \left(\frac{u \cdot v}{\|v\|^2}\right) v$
    - ❖Magnitude: $\|u\| \cos(\theta) = \frac{u \cdot v}{\|v\|}$, Direction: $\frac{v}{\|v\|}$



- Theorem) Every nonzero subspace of R^n has an orthonormal basis
  - Let W be a nonzero subspace of R^n, and let $\{w_1, \ldots, w_k\}$ be any basis for W. The following sequence steps will produce an orthogonal basis $\{v_1, \ldots, v_k\}$ for W

- Gram-Schmidt Process
  - Step 1) Let $v_1 = w_1$
  - Step 2) $v_2 = w_2 - proj_{W_1} w_2 = w_2 - \frac{w_2 \cdot v_1}{\|v_1\|^2} v_1$
    - ❖v_2 is orthogonal to v_1
    - ❖W_1: subspace spanned by v_1
  - Step 3) $v_3 = w_3 - proj_{W_2} w_3 = w_3 - \frac{w_3 \cdot v_1}{\|v_1\|^2} v_1 - \frac{w_3 \cdot v_2}{\|v_2\|^2} v_2$
    - ❖W_2: subspace spanned by v_1 and v_2





Source: Contemporary Linear Algebra

- 이런 그림은 어떻게 그릴까?

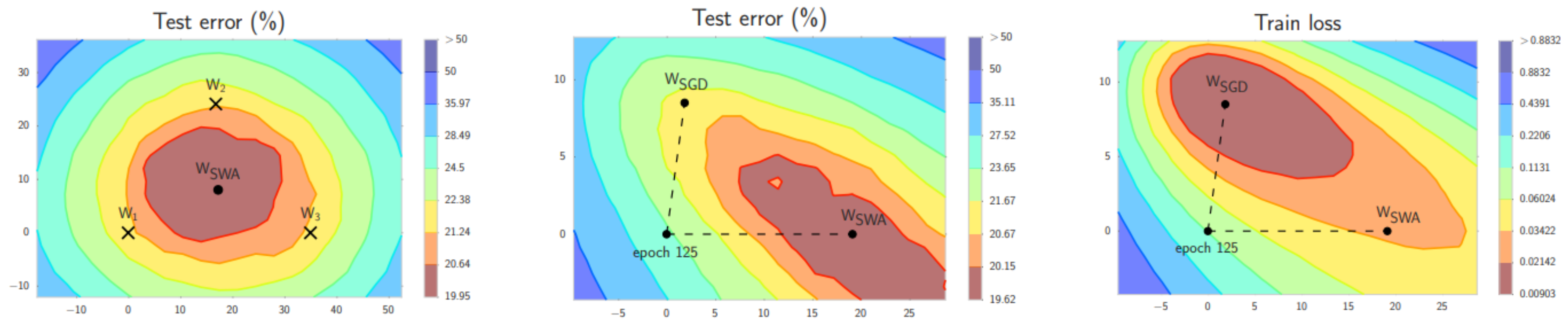


Figure 1: Illustrations of SWA and SGD with a Preactivation ResNet-164 on CIFAR-100[1]. **Left**: test error surface for three FGE samples and the corresponding SWA solution (averaging in weight space). **Middle** and **Right**: test error and train loss surfaces showing the weights proposed by SGD (at convergence) and SWA, starting from the same initialization of SGD after 125 training epochs.

[1] Suppose we have three weight vectors w_1, w_2, w_3. We set $u = (w_2 - w_1)$, $v = (w_3 - w_1) - \langle w_3 - w_1, w_2 - w_1 \rangle / \|w_2 - w_1\|^2 \cdot (w_2 - w_1)$. Then the normalized vectors $\hat{u} = u/\|u\|$, $\hat{v} = v/\|v\|$ form an orthonormal basis in the plane containing w_1, w_2, w_3. To visualize the loss in this plane, we define a Cartesian grid in the basis $\hat{u}, \hat{v}$ and evaluate the networks corresponding to each of the points in the grid. A point P with coordinates (x, y) in the plane would then be given by $P = w_1 + x \cdot \hat{u} + y \cdot \hat{v}$.

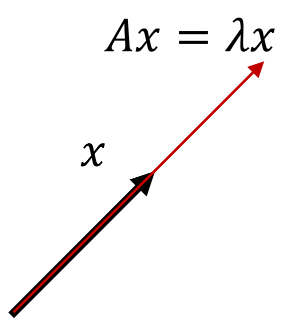
Source: https://arxiv.org/pdf/1803.05407.pdf

Linear Algebra - Decomposition

- Fixed points
  - All vectors in R^n that satisfy the equation $Ax = x, (I - A)x = 0$
  - They remain unchanged when multiplied by A
  - The vector $x = 0$ is a fixed point of every matrix A
- Note) If A is an $n \times n$ matrix, then the following statements are equivalent
  - A has nontrivial fixed points
  - $I - A$ is singular
  - $\det(I - A) = 0$
- 그렇다면, 다음의 식들을 만족하는 nonzero vectors 도 있을까?
  - $Ax = 2x, Ax = -3x, Ax = \sqrt{2}x,\ Ax = \lambda x,\ \lambda$: scalar

Source: Deep Learning Book, contemporary linear algebra

- **<u>Definition</u>**
  - If A is an $n \times n$ matrix, then a scalar λ is called an **eigenvalue** of A if there is nonzero vector x such that $Ax = \lambda x$. If λ is an eigenvalue of A, then every nonzero vector x such that $Ax = \lambda x$ is called an **eigenvector** of A corresponding to λ
  - $(\lambda I - A)x = 0$
- Note) If A is an $n \times n$ matrix, then the following statements are equivalent
  - λ is an eigenvalue of A
  - λ is a solution of the equation $\det(\lambda I - A) = 0$
  - The linear system $(\lambda I - A)x = 0$ has nontrivial solutions



Example

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

$\det(\lambda I - A) = (\lambda + 2)(\lambda - 5) = 0$

Eigenvalue: -2, 5

When $\lambda = -2$, $\begin{bmatrix} -3 & -3 \\ -4 & -4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = t\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

When $\lambda = 5$, $\begin{bmatrix} 4 & -3 \\ -4 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = t\begin{bmatrix} 3/4 \\ 1 \end{bmatrix}$

Source:

- Suppose that a matrix A has n linearly independent eigenvectors, $\{v^{(1)}, \dots, v^{(n)}\}$, with corresponding eigenvalues $\{\lambda_1, \dots, \lambda_n\}$.
- We may concatenate all of the eigenvectors to form a matrix V
  - $V = [v^{(1)}, \dots, v^{(n)}]$, one eigenvector per column
- We can concatenate the eigenvalues to form a vector $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_n]^T$
- $AV = \boldsymbol{\lambda} V$
- Eigendecomposition of A is then given by
  - $A = V diag(\boldsymbol{\lambda}) V^{-1}$
  - $diag(\boldsymbol{\lambda})$: $\boldsymbol{\lambda}$를 diagonal 로 하는 matrix
- 모든 행렬에 대해서 eigendecomposition 이 가능한것은 아닙니다.
- Every real symmetric matrix can be decomposed into eigenvalues and eigenvectors
  - $A = V diag(\boldsymbol{\lambda}) V^T$

Source:

- Every real symmetric matrix can be decomposed into eigenvalues and eigenvectors
  - $A = V diag(\boldsymbol{\lambda}) V^T$
- Quadratic expression form $f(x) = x^T A x$ s.t. $\|x\|_2 = 1$
  - For real symmetric A, f takes the value of eigenvalue whenever x is equal to an eigenvector of A
  - the maximum (minimum) value of f is the maximum (minimum) eigenvalue
- The matrix is singular if and only if any of the eigenvalues are zero
- A matrix whose eigenvalues are
  - all positive is called positive definite
  - all negative is called negative definite
  - All eigenvalues are negative or zero-valued is called negative semidefinite
  - All eigenvalues are positive or zero-valued is called negative semidefinite
    - $x^T A x \geq 0$

Source:

Singular Value Decomposition (SVD)

- The SVD allows us to discover some of the same kind of information as the eigendecomposition
- Every matrix has a singular value decomposition
- $A = UDV^T$
  - A: $m \times n$ matrix, U: $m \times m$ matrix, D: $m \times n$ matrix, V: $n \times n$ matrix
  - U and V are orthogonal matrix and D is diagonal matrix
  - U: left singular vectors A, the eigenvectors of AA^T
  - V: right singular vectors A, the eigenvectors of $A^T A$
- $AA^T = (UDV^T)(VD^T U^T) = U(DD^T)U^T$

Source:

- $Ax = y$
  - $x = By$
  - If A is taller than it is wide, then it is possible for this equation to have no solution
  - If A is wider than it is tall, then there could be multiple possible solutions

- Moore-Penrose pseudoinverse
  - Pseudoinverse of A is defined as a matrix
    - ❖ $A^+ = VD^+U^T$, where U, D, V are the SVD of A, and the D^+ is obtained by taking the reciprocal of its non-zero elements then taking the transpose of the resulting matrix
    - ❖ Note) $A = UDV^T$
- If A is taller than it is wide, A^+y is as close as possible to y in terms of Euclidean norm $||Ax - y||_2$
- If A is wider than it is tall, A^+y is the solution that has minimal Euclidean norm $||x||_2$

Source: