Saleh Sargolzaei
صالح سرگلزایی

# کورس یادگیری ماشین CS229

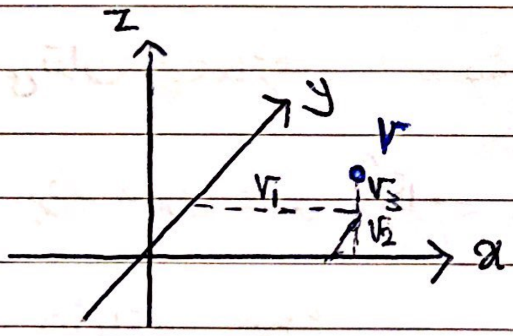
## مروری بر جبر خطی (Review of linear Algebra)

### اصطلاحات (Notation):

بردار (vector) - یک بردار مانند $v$، عضوی از $\mathbb{R}^d$ است.

جمله‌ی بالا به اینصورت نوشته می‌شود: $v \in \mathbb{R}^d$

$\mathbb{R}^d$ یک فضای حقیقی $d$-بعدی می‌باشد. به عنوان مثال:




$\longrightarrow$ نشان دهنده‌ی فضای $\mathbb{R}^3$ می‌باشد

وقتی می‌گوییم $v$ عضوی از فضای $\mathbb{R}^d$ است، به این معنا می‌باشد که $v$ یک نقطه در آن فضا می‌باشد. یک بردار به صورت زیر نمایش داده می‌شود:

* وکتور به صورت ستونی نمایش داده می‌شود:

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}$$

* در صورت محاسبه Transpose یک وکتور، آن را به صورت سطری نمایش می‌دهیم:

$$v^T = [v_1 \; v_2 \; \dots \; v_d]$$

ماتریس ( Matrix ) - برای ماتریس ها از حروف بزرگ استفاده

می شود. می گوییم : $A \in \mathbb{R}^{m \times n}$

به این معنا که A ماتریسی می باشد که m سطر و n ستون داشته و

اعداد داخل آن ، اعداد حقیقی می باشند.

$$A = \begin{array}{c c} & \\ m & \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \\ & n \end{array}$$

می توان این ماتریس را به چشم n تا بردار عضو $\mathbb{R}^m$ دید و یا m

بردار سطری عضو $\mathbb{R}^n$.

چند ماتریس مهم :

- Identity matrix : $(I \in \mathbb{R}^{n \times n})$ $\quad I = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & 1 \end{bmatrix}$

ماتریسی که قطر اصلی آن یک بوده است و بقیه ی مؤلفه های آن صفراند. به عبارت دیگر :

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

ماتریس Identity مربعی می باشد به این معنا که

تعداد سطرها و ستون های آن برابرند.

به ازای هر $A \in \mathbb{R}^{m\times n}$ ، $AI = A = IA$

دقت کنید که در سمت چپ تساوی $I \in \mathbb{R}^{n\times n}$ و در سمت راست $I \in \mathbb{R}^{m\times m}$

- **Diagonal matrix:**

$$D = \begin{bmatrix} \ddots & & 0 \\ & \ddots & \\ 0 & & \ddots \end{bmatrix}$$

$$D = \mathrm{diag}(d_1, d_2, \dots, d_n)$$

ماتریسی که همه‌ی مؤلفه‌هایی که در قطر اصلی نیستند برابر با صفر باشند. اعداد قطر اصلی می‌توانند هر مقداری را بگیرند. به عبارت دیگر:

$$D_{ij} = \begin{cases} d_i & i = j \\ 0 & i \neq j \end{cases}$$

- **Symmetric matrix:** ماتریس مربعی $A \in \mathbb{R}^{n\times n}$ متقارن است اگر $A = A^T$ . یعنی ماتریس و ترنسپوز آن برابر باشند.

* ترنسپوز (Transpose) یک ماتریس از جابه‌جایی سطرها و ستون‌های آن بدست می‌آید. یعنی $(A^T)_{ij} = A_{ji}$

ماتریس $A$ پادمتقارن (anti-symmetric) است اگر $A = -A^T$

**نکته -** برای هر ماتریس $A \in \mathbb{R}^{n\times n}$ ، ماتریس $A+A^T$ متقارن و ماتریس $A-A^T$ پادمتقارن می‌باشد. بنابراین می‌توان نشان داد که هر ماتریس $A$ در واقع حاصل جمع یک ماتریس متقارن و یک ماتریس پادمتقارن است:

$$A = \frac{1}{2}(A+A^T) + \frac{1}{2}(A-A^T)$$

**نکته -** برخی از ویژگی‌های Transpose عبارتند از:

$$(A^T)^T = A$$

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

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

**Trace -** trace یک ماتریس مربعی $A \in \mathbb{R}^{n\times n}$ ، که به صورت $tr(A)$ و یا گاهی $trA$ نمایش داده می‌شود، برابر است با حاصل جمع مؤلفه‌های قطر اصلی آن:

$$trA = \sum_{i=1}^{n} A_{ii}$$

Trace ویژگی‌هایی دارد که در صفحه‌ی بعد آمده اند:

- For $A \in \mathbb{R}^{n \times n}$, $trA = trA^T$
- For $A, B \in \mathbb{R}^{n \times n}$, $tr(A+B) = trA + trB$
- For $A \in \mathbb{R}^{n \times n}$, $t \in \mathbb{R}$, $tr(tA) = t\,trA$
- For $A, B$ s.t. $AB$ is Square, $trAB = trBA$
- For $A, B, C$ s.t. $ABC$ is Square, $trABC = trCAB = trBCA$

, and so on for the product of more matrices

## Vector - Vector Products:

با داشتن دو بردار $x$ و $y$ ، دو نوع ضرب می‌توان تعریف نمود:

① Inner Product/dot product : برای دو بردار $x, y \in \mathbb{R}^d$ ، ضرب داخلی دو بردار برابر است با:

$$[x_1 \; \cdots \; x_d] \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix} = \sum_{i=1}^{d} x_i y_i = x^T y$$

* برای انجام ضرب داخلی، دو بردار باید $d$ - بعدی باشند.

* ضرب داخلی دو بردار یک عدد حقیقی می‌باشد.

* $x^T y = y^T x$

② Outer Product : برای دو بردار $x \in \mathbb{R}^d$ و $y \in \mathbb{R}^p$ ،

برابر است با :

$$xy^T = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \begin{bmatrix} y_1 & \cdots & y_p \end{bmatrix} = \begin{bmatrix} x_1y_1 & \cdots & x_1y_p \\ & \ddots & \\ x_dy_1 & \cdots & x_dy_p \end{bmatrix}$$

$$yx^T = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_d \end{bmatrix} = \begin{bmatrix} y_1x_1 & \cdots & y_1x_d \\ \vdots & \ddots & \vdots \\ y_px_1 & \cdots & y_px_d \end{bmatrix}$$

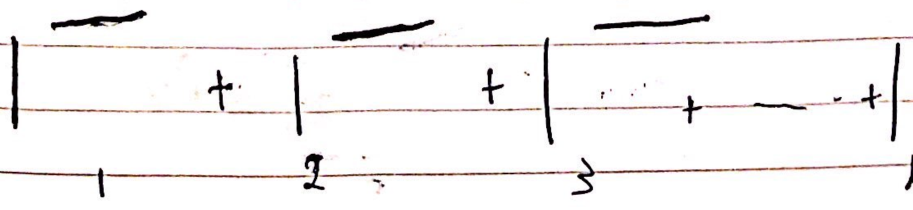
* در این حالت $x$ و $y$ می توانند اندازه های مختلفی داشته باشند

* حاصل یک ماتریس است : $yx^T \in \mathbb{R}^{p \times d}$ $xy^T \in \mathbb{R}^{d \times p}$

* همانطور که مشاهده می شود $xy^T \neq yx^T$

نکته - ماتریس حاصل از Outer Product ، Rank-1 است (در ادامه مفهوم Rank توضیح داده می شود)

نکته - حاصل جمع $k$ ماتریس Rank-1 ، در صورتی که بردارهای تشکیل دهنده‌ی آنها مستقل خطی (linearly Independent) باشند، برابر است با $\min(d,p,k)$ که در آن، بردارهای ستونی $d$ بعدی و بردارهای سطری $p$ بعدی اند :



$\text{rank} \leq \min(d,p,k)$

## Matrix - Vector Products :

حاصل ضرب ماتریس $A \in \mathbb{R}^{m \times n}$ و بردار $x \in \mathbb{R}^n$ ، بردار $y = Ax \in \mathbb{R}^m$

است.

ضرب ماتریس با یک بردار به طرق مختلف قابل تفسیری باشند:

1) اگر ماتریس A را به صورت مجموعه ای از بردارهای سطری در نظر بگیریم:

$$y = Ax = \begin{bmatrix} -- a_1^T -- \\ -- a_2^T -- \\ \vdots \\ -- a_m^T -- \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

به عبارت دیگر، مؤلفه‌ی i‌ام بردار y برابر با ضرب داخلی سطر i‌ام

ماتریس A با بردار x است: $y_i = a_i^T x$

2) از سوی دیگر، می‌توان ماتریس A را به صورت مجموعه ای از بردارهای ستونی

در نظر گرفت:

$$y = Ax = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} a^1 \end{bmatrix} x_1 + \begin{bmatrix} a^2 \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a^n \end{bmatrix} x_n$$

به عبارت دیگر، $y$ ترکیب خطی ستون‌های ماتریس $A$ می‌باشد، به طوریکه ضریب هر یک از ستون‌ها در این ترکیب خطی را مؤلفه‌های $x$ تعیین می‌کنند.

## Matrix - Matrix Products:

دو ماتریس $A \in \mathbb{R}^{m \times k}$ و $B \in \mathbb{R}^{k \times n}$ را در نظر بگیرید. راه‌های مختلفی برای تفسیر حاصل‌ضرب این دو ماتریس $C = AB \in \mathbb{R}^{m \times n}$ وجود دارد:

(1) در این تفسیر، ضرب دو ماتریس را به عنوان ضرب داخلی چند بردار در نظر می‌گیریم به طوریکه مؤلفه‌ی $(i,j)$ ماتریس $C$، حاصل‌ضرب داخلی $i$-امین سطر $A$ و $j$-امین ستون $B$ می‌باشد:

$$C = AB = \begin{bmatrix} — a_1^T — \\ — a_2^T — \\ \vdots \\ — a_m^T — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b^1 & b^2 & \cdots & b^n \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} a_1^T b^1 & \cdots & a_1^T b^n \\ \vdots & \ddots & \vdots \\ a_m^T b^1 & \cdots & a_m^T b^n \end{bmatrix}$$

دقت کنید که $a_i^T \in \mathbb{R}^k$ و $b_j \in \mathbb{R}^k$ ، بنابراین ضرب داخلی (inner product)

به درستی انجام می‌شود. این تفسیری است که با نمایش A به صورت سطری و

نمایش B به صورت ستونی حاصل می‌شود.

2 از طرف دیگر، با نمایش A به صورت ستونی و B به صورت سطری،

حاصل ضرب $C=AB$ را می‌توان به صورت حاصل جمع تعدادی outer product

در نظر گرفت:

$$C = AB = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^k \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & b_1^T & — \\ — & b_2^T & — \\ & \vdots & \\ — & b_k^T & — \end{bmatrix}.$$

$$= \sum_{i=1}^{k} a^i b_i^T$$

$$= \begin{matrix} | \\ a^1 \\ | \end{matrix} \begin{matrix} — b_1^T — \end{matrix} + \begin{matrix} | \\ a^2 \\ | \end{matrix} \begin{matrix} — b_2^T — \end{matrix} + \cdots + \begin{matrix} | \\ a^k \\ | \end{matrix} \begin{matrix} — b_k^T — \end{matrix}$$

$$\text{rank}(C) \leq \min(m, n, k) \quad *$$

ج) با نمایش A به صورت ستونی و B به صورت ستونی داریم:

$$C = AB = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^k \\ | & | & & | \end{bmatrix} \begin{bmatrix} | & | & & | \\ b^1 & b^2 & \cdots & b^n \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{k} a^i b^1_i & \cdots & \sum_{i=1}^{k} a^i b^n_i \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ Ab^1 & Ab^2 & \cdots & Ab^n \\ | & | & & | \end{bmatrix}$$

**نکته-** برخی از خواص پایه ضرب ماتریس ها عبارتند از:

- associative ( شرکت پذیری ) : $(AB)C = A(BC)$
- distributive ( توزیع پذیری ) : $A(B+C) = AB + AC$
- ضرب ماتریسی در حالت کلی جابه‌جایی پذیر ( commutative ) نمی باشد : $AB \neq BA$

# Norms:

norm یک بردار، در تعریف غیررسمی، معیاری از "طول" بردار می‌باشد.
(اندازه)

به عنوان مثال یکی از پر استفاده ترین norm ها، اقلیدسی (Euclidean)

یا $l_2$-norm است:
$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

* نرم یک بردار $x$ با علامت $\|x\|$ نمایش داده می‌شود.

* دقت کنید که $\|x\|_2^2 = x^T x$

در تعریف رسمی، یک norm هر تابعی است $f: \mathbb{R}^n \to \mathbb{R}$

که 4 خاصیت زیر را داشته باشد:

1) برای هر $x \in \mathbb{R}^n$، $f(x) \geq 0$ (non-negativity)

2) $f(x)=0$ اگر و فقط اگر $x=0$ (definiteness)

3) برای هر $x \in \mathbb{R}^n$ و $t \in \mathbb{R}$، $f(tx)=|t|f(x)$ (homogeneity)

4) برای هر $x, y \in \mathbb{R}^n$، $f(x+y) \leq f(x)+f(y)$ (triangle inequality)

مثال های دیگری از norm ها عبارتند از :

$l_1$ norm : $\|x\|_1 = \sum_{i=1}^{n} |x_i|$

$l_\infty$ norm: $\|x\|_\infty = \max_i |x_i|$

* تمامی سه norm بحث شده تا اینجا مثال هایی از خانواده $l_p$ norms هستند،

که شامل پارامتر حقیقی $1 \le p$ بوده و به صورت زیر تعریف می‌شود:

$$\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

نکته ـ نرم ها می‌توانند برای ماتریس ها نیز تعریف شوند. مانند

Frobenius norm:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2}$$

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

## Linear Independence and Rank:

مجموعه‌ای از بردارها $\{x_1, x_2, ..., x_n\} \in \mathbb{R}^m$، مستقل خطی یا linearly independent می باشند اگر هیچ یک از بردارها ترکیبی خطی از سایر بردارها نباشد. متقابلاً، اگر بتوان برداری در این مجموعه را به صورت ترکیبی خطی از سایر بردارهای موجود در مجموعه نمایش داد، بردارها وابسته خطی یا linearly dependent می باشند. $x_n$ ترکیبی خطی از سایر بردارها است

اگر:
$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

برای مقادیر عددی $\alpha_1, \alpha_2, ..., \alpha_{n-1} \in \mathbb{R}$.

برای مثال، بردارهای:

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

وابسته خطی یا linearly dependent اند زیرا: $x_3 = -2x_1 + x_2$

* Column Rank (رتبه ستونی) یک ماتریس $A \in \mathbb{R}^{m \times n}$، برابر است با اندازه‌ی بزرگترین زیرمجموعه از ستون های ماتریس که تشکیل یک مجموعه مستقل خطی یا linearly Independent را می دهند.

Scanned with CamScanner

* به طور مشابه، row rank (رتبه سطری) یک ماتریس A، برابر است

با بیشترین تعداد سطرهای A که یک مجموعه مستقل خطی یا linearly independent

را تشکیل می دهند.

**نکته** - برای هر ماتریس $A \in \mathbb{R}^{m \times n}$، قابل اثبات است که column rank

و row rank برابرند. به همین دلیل هر دو مقدار را rank (رتبه)

ماتریس می نامیم و به صورت $rank(A)$ نمایش می دهیم.

موارد زیر، برخی از خواص rank (رتبه) می باشند:

- برای $A \in \mathbb{R}^{m \times n}$، $rank(A) \leq min(m,n)$. اگر $rank(A) = min(m,n)$،

ماتریس full rank (رتبه کامل) نامیده می شود

- برای $A \in \mathbb{R}^{m \times n}$، $rank(A) = rank(A^T)$
- برای $A \in \mathbb{R}^{m \times n}$ و $B \in \mathbb{R}^{n \times p}$، $rank(AB) \leq min(rank(A), rank(B))$
- برای $A, B \in \mathbb{R}^{m \times n}$، $rank(A+B) \leq rank(A) + rank(B)$

# The Inverse of a Square Matrix
(معکوس یک ماتریس مربعی)

معکوس (Inverse) یک ماتریس مربعی $A \in \mathbb{R}^{n \times n}$ را با $A^{-1}$ نشان می دهیم. ماتریس معکوس $A^{-1}$، یک ماتریس یکتا بوده به طوریکه:

$$A^{-1}A = I = AA^{-1}$$

نکته- همه‌ی ماتریس ها، معکوس پذیر نمی باشند. به عنوان مثال ماتریس های غیر مربعی، معکوس ندارند. با این وجود، ماتریس های مربعی ای هم وجود دارند که معکوس ندارند.

برای ماتریس مربعی $A$، می گوییم Invertible (معکوس پذیر) یا non-singular (غیر منفرد) است، اگر $A^{-1}$ وجود داشته باشد. در مقابل $A$ را non-invertible (معکوس ناپذیر) یا Singular (منفرد) می نامیم، اگر $A^{-1}$ وجود نداشته باشد.

* برای اینکه ماتریس مربعی $A \in \mathbb{R}^{n \times n}$، معکوسی مانند $A^{-1}$ داشته باشد، باید $A$، full rank باشد یعنی $rank(A)=n$.

برخی از ویژگی های $^{-1}$ (معکوس) عبارتند از (فرض اینکه ماتریس های $A, B \in \mathbb{R}^{n \times n}$ ، non-singular باشند) :

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^{-1})^{T} = (A^{T})^{-1}$ (به همین دلیل گاهی به صورت $A^{-T}$ نشان داده می شود)

## Column Space, Row Space and Nullspace

توضیح "Span" : span مجموعه‌ای از بردارها $\{x_1, x_2, \ldots, x_n\}$ ، مجموعه‌ای است شامل تمامی بردارهایی که می‌توان به صورت ترکیبی خطی از $\{x_1, x_2, \ldots, x_n\}$ نشان داده شوند. به عبارت دیگر :

$$span(\{x_1, x_2, \ldots, x_n\}) = \left\{ v : v = \sum_{i=1}^{n} \alpha_i x_i \ , \ \alpha_i \in \mathbb{R} \right\}$$

نکته- می‌توان نشان داد که اگر مجموعه $\{x_1, x_2, \ldots, x_n\}$ شامل $n$ بردار مستقل خطی (linearly independent) باشد، به طوری‌که $x_i \in \mathbb{R}^n$، آنگاه $\text{span}(\{x_1, \ldots, x_n\}) = \mathbb{R}^n$. به عبارت دیگر، هر بردار $v \in \mathbb{R}^n$ را می‌توان به صورت ترکیب خطی از بردارهای $x_1$ تا $x_n$ نوشت.

Column Space یا range یک ماتریس $A \in \mathbb{R}^{m \times n}$، با $\mathcal{R}(A)$ نمایش داده می‌شود، عبارت است از span مجموعه بردارهای ستون‌های ماتریس A (به عبارت دیگر، مجموعه تمام بردارهایی که از ترکیب خطی ستون‌های A به دست می‌آید).

به طور مشابه، row space ماتریس A، عبارت است از span مجموعه بردارهای سطرهای ماتریس A.

Nullspace ماتریس $A \in \mathbb{R}^{m \times n}$، با $\mathcal{N}(A)$ نمایش داده می‌شود و مجموعه تمام بردارهایی می‌باشد که در صورت ضرب با ماتریس A، حاصل برابر با بردار $0$ می‌شود. به عبارت دیگر:

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$$

# Geometrical Interpretation of Matrix_Vector Product
( تفسیر هندسی ضرب ماتریس و بردار )

تا به اینجا، روش های مختلف برای تفسیر ضرب ماتریس $A \in \mathbb{R}^{m\times n}$ و بردار $x \in \mathbb{R}^n$ را دیدیم، که حاصل آن برابر است با بردار $y = Ax \in \mathbb{R}^m$.

با این حال، راه بهتری برای درک این فرآیند وجود دارد. می توان به جای فکر کردن به ضرب اعداد داخل ماتریس و بردار، ماتریس را به چشم "تابعی" دید که ورودی آن بردار $x \in \mathbb{R}^n$ بوده و خروجی آن بردار $y \in \mathbb{R}^m$ می باشد.

به عنوان مثال، ماتریس مربعی و full rank $A \in \mathbb{R}^{3\times 3}$ را در نظر بگیرید ( مربعی بودن و ابعاد ماتریس جهت سادگی نمایش انتخاب شده اند). در صورتی که این ماتریس را به عنوان "تابعی" به شکل $A(x) = y$ در نظر بگیرید، می توان دید که این تابع هر برداری را در فضای ورودی ( Input space $\mathbb{R}^3$ ) به برداری یکتا در فضای خروجی ( Output space $\mathbb{R}^3$ )، نگاشت می نماید.

* مواردی که در دو خط بالا زیر آنها خط کشیده شده است، یعنی توانایی نگاشت هر برداری در فضای $\mathbb{R}^3$ به برداری یکتا در فضای خروجی $\mathbb{R}^3$، نتیجه ی full rank

بودن ماتریس A می باشد. در ادامه درباره‌ی حالتی که ماتریس rank deficient

(دارای کمبود رتبه) باشد، صحبت می شود.

علاوه بر A، به دلیل مربعی بودن و full rank بودن A، ماتریس $A^{-1}$ نیز

وجود دارد. $A^{-1}$ نیز مانند تابعی عمل می کند که هر نقطه‌ای را در فضای

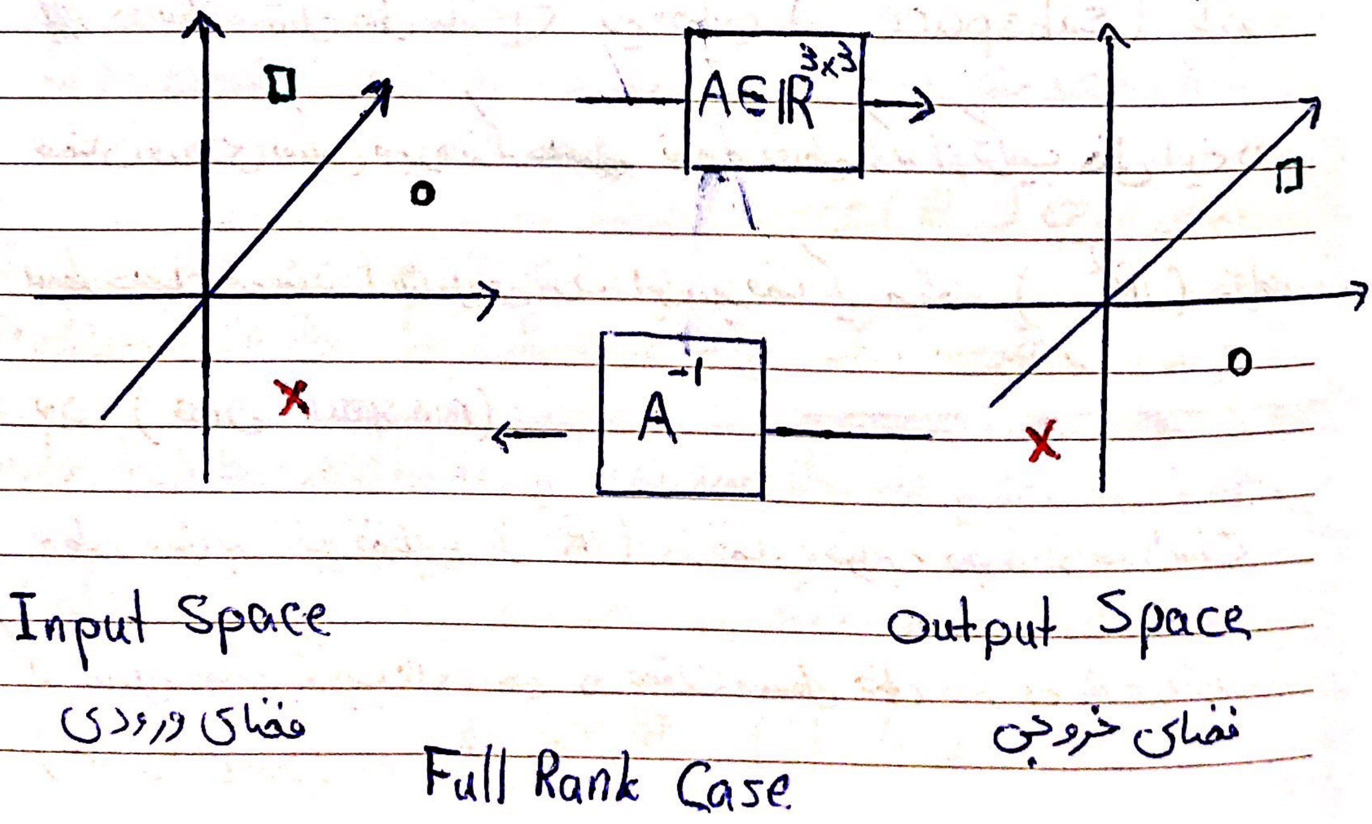
خروجی A، به نقطه‌ی مبدأ آن در فضای ورودی A نگاشت می نماید. به

عبارت دیگر، معکوس عملیات A را انجام می دهد. شکل زیر، درک بهتری از مباحث

بررسی شده می دهد:

* در شکل زیر نقاط (بردار)های مختلف در فضای ورودی و نگاشت یکتا آنها در فضای

خروجی، با اشکال مختلف نمایش داده شده اند.




Full Rank Case

* اگر A، full rank نباشد ( rank deficient )، چه اتفاقی می افتد؟!

فرض کنید ماتریس A یک سطر داشته باشد که از ترکیب خطی دو سطر دیگر حاصل شده باشد ( و از آنجایی که row rank = column rank، یکی از ستون ها نیز ترکیب خطی دو ستون دیگر خواهد بود ). به عبارت دیگر برای ماتریس مربعی $A \in \mathbb{R}^{3\times3}$، $rank(A) = 2$.

در اینصورت، برخلاف حالت full rank، column space و row space ماتریس A کل فضای $\mathbb{R}^3$ را شامل نمی شود.

دو سطر مستقل خطی ماتریس زیرفضایی ( subspace ) را در فضای ورودی تشکیل می دهند که تمامی نقاطی را که از ترکیب خطی این دو سطر حاصل می شوند را شامل می شود. این زیرفضا یک صفحه ( $\mathbb{R}^2$ ) خواهد بود. ( همان row space )

به طور مشابه زیرفضایی ( $\mathbb{R}^2$ ) در فضای خروجی وجود خواهد داشت که شامل تمامی ترکیب های خطی دو ستون مستقل خطی ماتریس A می شود. ( همان column space )

**نکته** - برخی از مفاهیم صفحه‌ی قبل، نیاز به درک مفاهیم زیر دارد:

vector space (فضای برداری): یک فضای برداری، مجموعه‌ای از بردارها است که نسبت به عمل ترکیب خطی بسته باشد (closed under linear combinations)

به عبارت دیگر، در صورت محاسبه ترکیب خطی از هر تعداد اعضای این مجموعه، حاصل همچنان در مجموعه باشد.

subspace (زیرفضا): یک زیرفضا، فضای برداری است که در فضای برداری دیگری قرار گرفته باشد. دقت کنید که یک زیرفضا می‌تواند با فضایی که در آن قرار دارد برابر باشد، مانند حالت full rank در دو صفحه‌ی قبل که row space شامل تمامی فضای ورودی ($\mathbb{R}^3$) و column space شامل تمامی فضای خروجی ($\mathbb{R}^3$) می‌شد.

* بر اساس تعریف، یک زیرفضا باید همیشه از مبدأ (بردار صفر) عبور کند، زیرا همیشه می‌توان ضرایب ترکیب خطی مورد نظر را برابر با صفر قرار داد و حاصل، که بردار صفر خواهد بود، باید در مجموعه باشد. بر اساس این نکته، کوچکترین زیرفضا ممکن خود بردار صفر (مبدأ)، $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ می‌باشد.

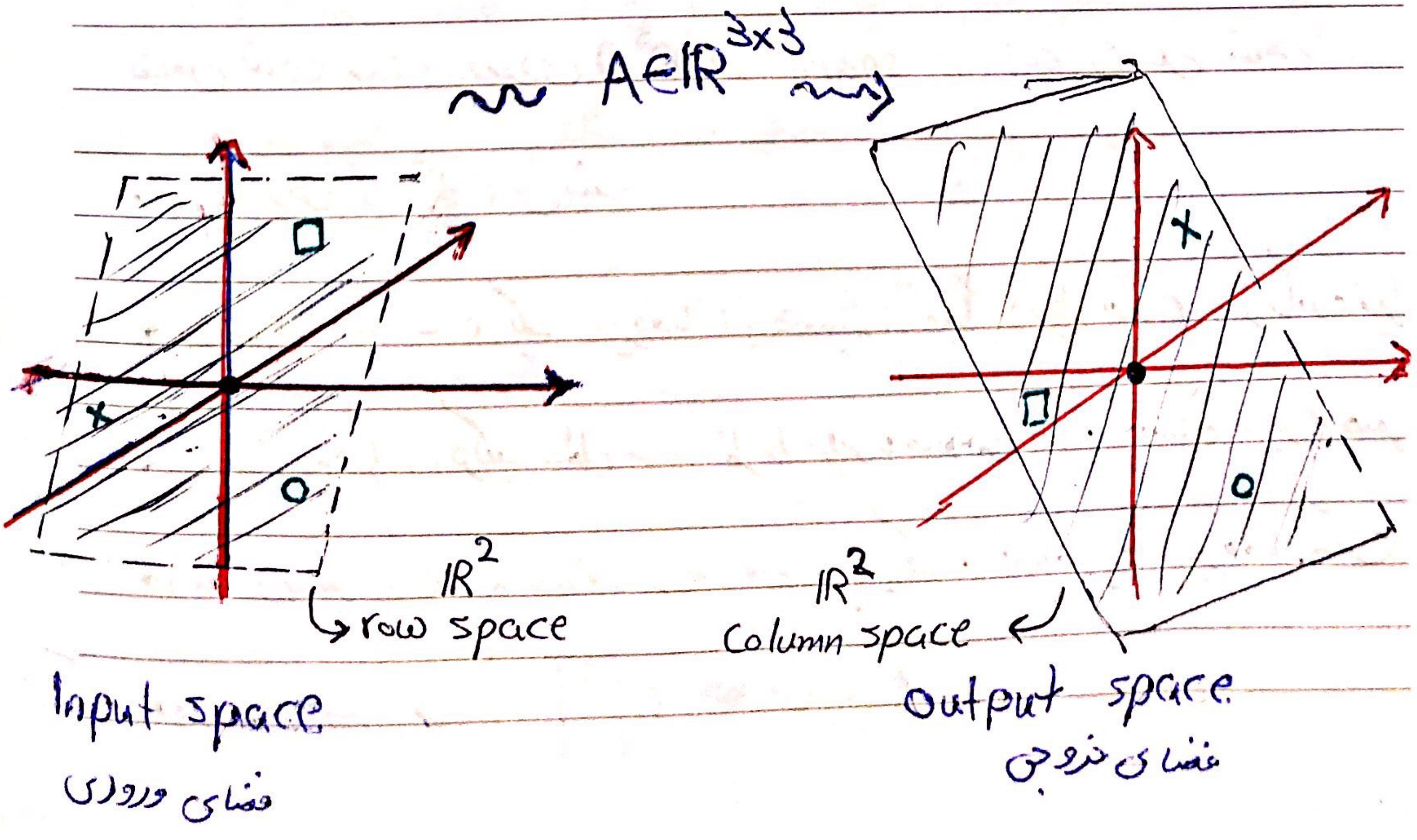
Scanned with CamScanner

با توجه به تعریف زیرفضا، زیرفضاهای فضای $\mathbb{R}^3$ عبارتند از:

- مبدأ / origin $(\mathbb{R}^0)$
- $(\mathbb{R}^1)$ هر خطی که از مبدأ عبور کند
- $(\mathbb{R}^2)$ هر صفحه‌ای که از مبدأ عبور کند
- $(\mathbb{R}^3)$ کل فضا

---

بنابراین برای $A \in \mathbb{R}^{3\times3}$ و $\text{rank}(A)=2$، هر نقطه روی زیرفضای حاصل از row space ماتریس، به نقطه‌ای یکتا بر روی زیرفضای حاصل از column space ماتریس نگاشت می‌گردد. توضیحات در شکل زیر نمایش داده شده اند:



با توجه به شکل می‌توان دید که از آنجایی که هر ترکیب خطی از ستون های ماتریس A

روی زیرفضای موجود در فضای خروجی قرار می‌گیرند، نقاطی (بردارهایی) که در

فضای خروجی قرار داشته اما عضوی از زیرفضای $\mathbb{R}^2$ نیستند، غیرقابل دسترسی

هستند. به عبارت دیگر هیچ برداری مانند $x \in \mathbb{R}^3$ وجود ندارد به طوریکه

حاصل $Ax$، برداری خارج از زیرفضا باشد. ( $Ax \in$ Column space ، برای همه $x$ )

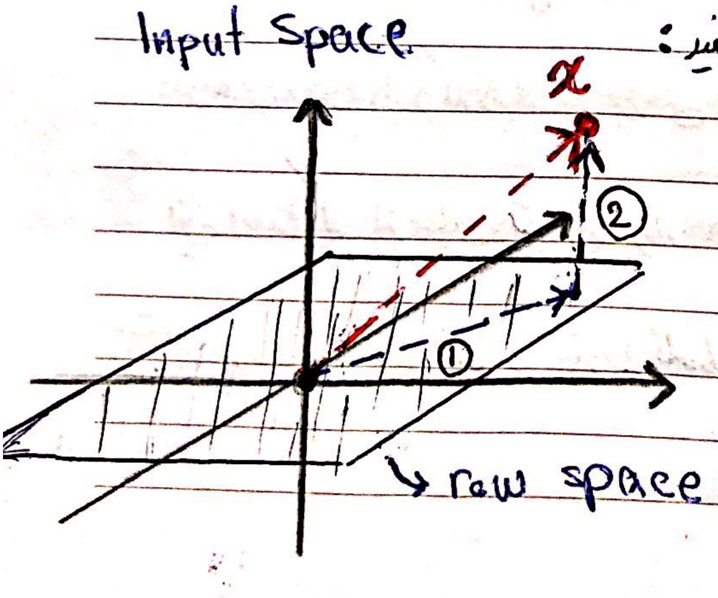
از طرف دیگر، می‌توان دید که هر بردار $x \in \mathbb{R}^3$ در فضای ورودی را می‌توان

با ماتریس A ضرب نمود (به عنوان ورودی تابع A). پس سؤالی که پیش

می‌آید این است که نقاطی که بر روی row space قرار ندارند، در صورت

ضرب شدن با A، به چه نقطه‌ای در فضای خروجی نگاشت می‌شوند؟!

برای درک جواب این سؤال به شکل زیر دقت کنید:




① Proj($x$ ; row space)

② Proj($x$ ; nullspace)

$$\vec{x} = ① + ②$$

می‌توان دید که هر برداری مانند $x$ در شکل، که خارج از rowspace(A)

می باشد را می‌توان به صورت حاصل جمع دو بردار دیگر نوشت:

① برداری که عضوی از rowspace(A) است و نزدیک ترین فاصله

اقلیدسی را به $x$ دارد [همان $Proj(x; rowspace)$، یعنی تصویر بردار $x$ بر روی

زیرفضای حاصل از rowspace(A)].

② برداری که عمود بر rowspace(A) بوده و عضوی از $N(A)$ می باشد.

[همان $Proj(x, nullspace)$، یعنی تصویر بردار $x$ بروی $N(A)$.

تعریف Nullspace در بخش های قبل آمده است.]

منظور از projection بردار $x$ بر روی یک زیرفضا (subspace)

چیست؟ برای ~~projection~~ محاسبه‌ی projection بردار $x$ بر روی یک زیرفضا

فرمولی وجود دارد که در ادامه بررسی می‌شود، اما مفهوم $Proj(x; subspace)$

این است که از میان تمامی بردارهای $v \in subspace$، کدام یک نزدیک ترین

به بردار $x$ هست (کمترین فاصله اقلیدسی را با $x$ دارد). و یا به صورت

ریاضیاتی :

$$Proj(x; subspace) = \underset{v \in subspace}{argmin} \|x - v\|_2$$

مشخص است که نزدیک ترین نقطه (بردار)، نقطه‌ی تلاقی زیرفضا و خطی است

که از بردار $x$ به زیرفضا عمود می‌شود.

بنابراین با نوشتن $x$ به صورت جمع دو بردار خواهیم داشت (برای سادگی فرمول‌ها

$Proj(x; row space)$ با $x_R$ و $Proj(x; Nullspace)$ با $x_N$ نمایش داده شده):

$$Ax = A(x_R + x_N)$$

$$= Ax_R + \underbrace{Ax_N}_{0}$$

$$= Ax_R$$

می‌دانیم که $x_N$ در $N(A)$ می‌باشد و بر اساس

تعریف Nullspace برای هر بردار $v \in N(A)$

داریم $Av = 0$

یعنی هر نقطه‌ای خارج از row space(A)، در صورت ضرب شدن با A،

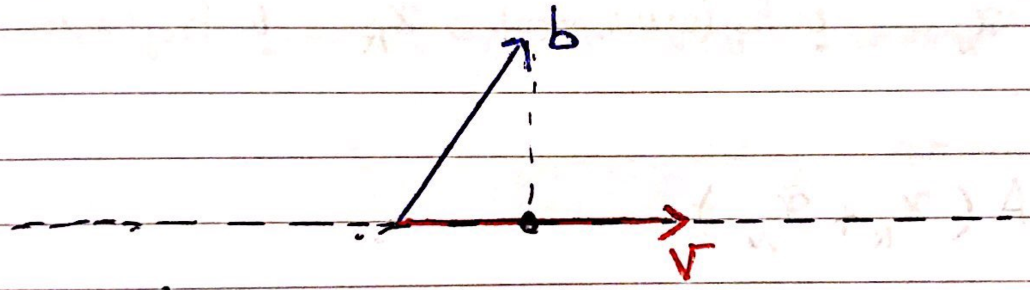
بر روی row space(A)، project شده و سپس به نقطه‌ای در

فضای خروجی نگاشت می‌شود که بردار حاصل از Projection بر روی آن نگاشت می‌شود.

# Projection

تا اینجا با مفهوم projection آشنا شدیم، اما چگونه projection محاسبه می‌گردد؟!

حالت ساده‌ای را در نظر بگیرید که دو بردار $v$ و $b$ وجود داشته و می‌خواهیم بردار $b$ را بر روی زیرفضای یک‌بعدی که $v$ عضو آن است project کنید:




دقت کنید که با در نظر گرفتن تمامی ضرایب $v$، می‌توان خطی را تشکیل داد که در واقع همان زیرفضایی می‌باشد که $v$ عضوی از آن است ( $\mathbb{R}^1$ ) .

در اینصورت برای $v$ ماتریسی به نام projection matrix وجود دارد که در صورت ضرب شدن با هر برداری در فضا، حاصل ضرب برابر خواهد بود با:

projection بردار ضرب شده بر روی زیرفضایی که $v$ عضو آن است:

$$\text{Projection Matrix}(v) = \left[ \frac{v v^T}{v^T v} \right]$$

با کمی بررسی فرمول می‌توان فهمید چرا $\left[\frac{vv^T}{v^Tv}\right] b$ ، projection بردار b را بر روی زیرفضای v می‌دهد :

$$\frac{v\,v^T}{v^T\,v} b = \frac{v\,v^T}{\|v\|_2^2} b = \frac{v}{\|v\|_2} \cdot \frac{v^T}{\|v\|_2} \cdot b$$

می‌دانیم که با تقسیم v بر طول آن (l$_2$-norm) ، برداری حاصل می‌شود که طول آن برابر است با یک ( $\|\tilde{v}\|_2 = 1$ ) و در جهت بردار v است.

* دقت کنید به بردار $x \in \mathbb{R}^n$ ، normalized (نرمال شده) گفته می‌شود اگر $\|x\|_2 = 1$ .

در ادامه v را با نسخه‌ی نرمال شده اش جایگزین می‌کنیم و آنرا $\tilde{v}$ می‌نامیم :

$$\frac{v\,v^T}{v^T v} b = \tilde{v}\,\tilde{v}^T b$$

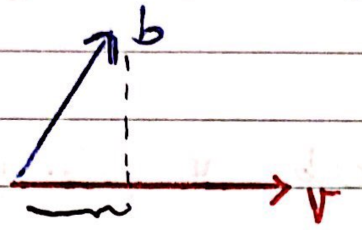
* با توجه به این شکل از فرمول می‌توان دید که هر برداری در زیرفضایی که v عضو آن است ، در نهایت به projection matrix مشابهی ختم می‌شود زیرا تمامی آنها نرمال شده و $\tilde{v}$ حاصل می‌گردد.

Scanned with CamScanner

با پرانتز گذاری فرمول قبل داریم: $\tilde{v}(\tilde{v}^T b)$

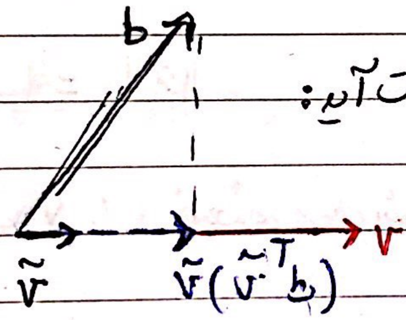
می دانیم که $\tilde{v}^T b$ در واقع ضرب داخلی برداری به طول یک ($\tilde{v}$) با برداری دیگر ($b$) است. ضرب داخلی یک بردار به طول یک با برداری دیگر، بزرگی (طول projection) بردار را در جهت بردار یکه می دهد:

← یا بردار یکه



$\tilde{v}^T b$ این اندازه را می دهد ←

در نهایت $\tilde{v}^T b$ با بردار یکه $\tilde{v}$ ضرب می شود تا آنرا در جهت $v$ تغییر اندازه داده و تصویر نهایی $b$ بر روی زیرفضا بدست آید:




---

چه اتفاقی می افتد اگر زیرفضایی که می خواهیم بر روی آن project کنیم از ترکیب خطی چندین بردار مستقل خطی حاصل شود (زیرفضا بزرگتر از $\mathbb{R}^1$ است).

در اینصورت بجای بردار $v$، ماتریسی مانند $X$ خواهیم داشت که ستون های

آن بردارهایی هستند که زیرفضای مورد نظر را تشکیل می دهند. به عنوان مثال در

تفسیر هندسی ضرب ماتریسی $V$ دیدیم که در صورتیکه برداری خارج از row space(A) و بردار $x$

باشد، حاصل ضرب برابر با حاصل ضرب ماتریس با Proj(x; row space) است.

در اینصورت به جای ماتریس $V$ را قرار می دهیم $X = A^T$. دلیل استفاده

از ترنسپوز این است که می خواهیم بر روی زیرفضایی که سطرها تشکیل می دهند تصویر کنیم،

بنابراین سطرها را در ستون های $X$ قرار داده ایم.

به عنوان مثال برای تصویر $b$ بر روی زیرفضایی برابر با $\mathbb{R}^3$ ، $X$ شکلی به صورت زیر

خواهد داشت:

$$X = \begin{bmatrix} | & | & | \end{bmatrix}$$

بنابراین فرمول projection matrix به صورت زیر در می آید:

Projection matrix for projecting on R(X) $= X(X^TX)^{-1}X^T$

* دقت کنید که $V^TV$ یک عدد بود بنابراین می توانستیم عمل تقسیم را انجام دهیم.

اما $X^TX$ ماتریس است و باید به جای تقسیم در $(X^TX)^{-1}$ ضرب کنیم.

# Eigenvalues and Eigenvectors
(مقادیر ویژه و بردارهای ویژه)

برای ماتریس مربعی $A \in \mathbb{R}^{n \times n}$ ، می‌گوییم $\lambda \in \mathbb{C}$ یک مقدار ویژه (eigenvalue) ماتریس A و $x \in \mathbb{C}^n$ بردار ویژه (eigenvector) مربوطه است اگر:

$$Ax = \lambda x \quad , \quad x \neq 0$$

به این معنا که با ضرب ماتریس A در بردار ویژه $x$ ، برداری حاصل می‌گردد که در جهت بردار $x$ بوده اما طول آن با ضریب $\lambda$ تغییر کرده است.

* دقت کنید که در این تعریف $\lambda$ و مقادیر $x$ ، اعداد مختلط (complex numbers) می‌باشند. در این مورد دلیل این امر بررسی نمی‌شود اما در ادامه حالت خاصی معرفی می‌گردد که بسیار کاربردی بوده و در آن تنها با اعداد حقیقی سروکار خواهیم داشت.

نکته- برای هر مقدار $c \in \mathbb{C}$ و بردار ویژه $x \in \mathbb{C}^n$ داریم:

$$A(cx) = c(Ax) = c(\lambda x) = \lambda(cx)$$

بنابراین $cx$ نیز یک بردار ویژه با مقدار ویژه مشابه با $x$ خواهد بود.

به همین دلیل معمولاً زمانیکه درباره‌ی بردار ویژه مربوط به مقدار ویژه $\lambda$ صحبت می‌کنیم،

فرض می‌کنیم که بردار ویژه مربوطه نرمالسازی شده است تا طول یک داشته باشد. دقت

کنید که حتی با وجود این فرض مقداری ابهام وجود خواهد داشت، زیرا $x$ و $-x$

هر دو میتوانند بردار ویژه مربوطه باشند!!

## Eigenvalues and Eigenvectors of Symmetric Matrices

( مقادیر ویژه و بردارهای ویژه ماتریس های متقارن )

در یادگیری ماشین، حالت های زیادی وجود دارد که با ماتریس‌های مربعی متقارن

سروکار داریم که مقادیر حقیقی دارند. براساس تئوری spectral،

برای ماتریس $A \in \mathbb{R}^{n\times n}$ و $A = A^T$ (symmetric)، در

خصوص مقادیر ویژه و بردارهای ویژه آن، دو ویژگی زیر وجود دارد:

① همه‌ی مقادیر ویژه‌ی آن ( $\lambda_1, \lambda_2, \dots, \lambda_n$ )، اعداد حقیقی هستند

② بردارهای ویژه آن نسبت به هم orthonormal هستند

نکته - توضیح بردارهای orthogonal (متعامد): دو بردار $x, y \in \mathbb{R}^n$،

orthogonal هستند اگر $x^T y = 0$. بنابراین اگر دو بردار norm غیر صفر

داشته باشند، به این معنا است که آنها نسبت به هم در زاویه $90^\circ$ قرار دارند.

در فضای $\mathbb{R}^n$، نهایتاً $n$ بردار با norm غیر صفر می توانند دو به دو orthogonal باشند.

• توضیح بردارهای orthonormal: اگر بردارها علاوه بر orthogonal (متعامد) بودن، norm برابر با یک (نرمالسازی شده) نیز داشته باشند، به آنها orthonormal گوییم.

• توضیح ماتریس orthogonal: ماتریس مربعی $U \in \mathbb{R}^{n \times n}$ را orthogonal می نامیم اگر همه ی ستون ها (همچنین همه سطرها) نسبت به هم orthonormal باشند. بنابراین بر اساس تعریف متعامد بودن دو بردار داریم:

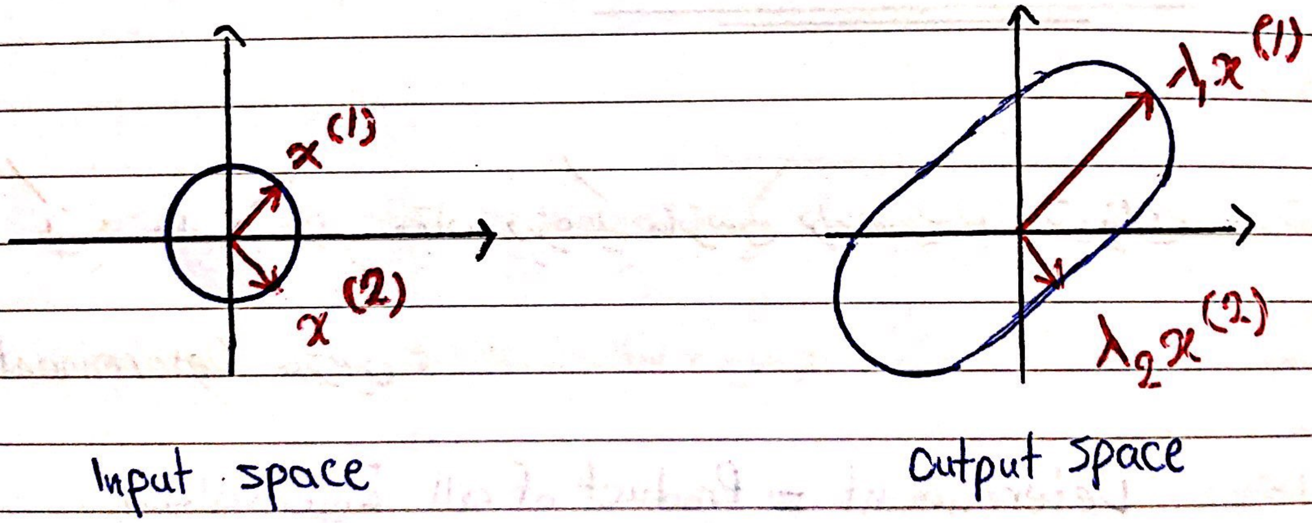
$$U^T U = I = U U^T$$

پس می توان نتیجه گرفت که برای یک ماتریس orthogonal:

$$U^T = U^{-1}$$

## تفسیر هندسی Eigevectors و Eigenvalues:

ماتریس مربعی و متقارن $A \in \mathbb{R}^{2\times 2}$ را در نظر بگیرید. با ضرب تمامی نقاطی که به شعاع واحد از مبدأ فضای ورودی A قرار دارند با ماتریس A (نقاطی که روی سطح دایره‌ی شکل زیر می‌باشند)، بیضی‌ای در فضای خروجی A حاصل می‌گردد:




همانطور که در شکل مشاهده می‌شود، دو برداری که در فضای ورودی نمایش داده شده‌اند، در صورت ضرب با ماتریس A تنها به اندازه‌ی $\lambda_i$ تغییر اندازه داده اند. این بردارها $\{x^{(1)}, x^{(2)}\}$ بردارهای ویژه A و $\{\lambda_1, \lambda_2\}$ مقادیر ویژه متناظر می‌باشند. این در حالی است که نقاط دیگر بر روی دایره، علاوه بر تغییر اندازه، تغییر جهت نیز می‌دهند.

نکته: در صورتیکه یکی از مقادیر ویژه برابر با صفر باشد (به عنوان مثال $\lambda_2$)، شکل حاصل در فضای خروجی فشرده گردیده و تبدیل به یک خط می شود. در اینصورت ماتریس Rank Deficient است.

بنابراین در تعریف دیگری از Rank یک ماتریس می توان گفت:

Rank ماتریس برابر است با تعداد مقادیر ویژه مخالف صفر.

نکته: حاصل ضرب مقادیر ویژه یک ماتریس برابر است با دترمینان Determinant ماتریس:

$$\text{Determinant} = \text{Product of all Eigenvalues}$$

و از این تفسیر می توان نتیجه گرفت:

$$\text{Determinant} = \frac{\text{Volume (Output shape)}}{\text{Volume (Input shape)}}$$

بنابراین اگر ماتریس Rank Deficient باشد به این معنی است که دارای مقادیر ویژه صفر بوده و دترمینان برابر با صفر است.

بنابراین دترمینان درکی از میزان گسترش یا فشرده شدن فضا توسط ماتریس

را می دهد.

همچنین از تفسیر هندسی مشخص است که در صورت صفر بودن مقادیر ویژه،

که صفر بودن دترمینان را نیز به همراه دارد، ماتریس معکوس ناپذیر است زیرا

با فشرده شدن دایره به یک خط، عملاً امکان معکوس کردن بُعد فشرده شده

از دست می رود.

نکته - به مجموعه ی مقادیر ویژه یک ماتریس، Spectrum ماتریس

گویند. Spectrum یک ماتریس، اطلاعات زیادی از عملیاتی که ماتریس

انجام می دهد به ما می دهد.

---

## Quadratic Forms

با داشتن ماتریس مربعی $A \in \mathbb{R}^{d \times d}$ و بردار $x \in \mathbb{R}^d$، مقدار

حقیقی $x^T A x$، quadratic form نامیده می شود.

در حالت کلی، هنگام کار با quadratic form فرض می شود که ماتریس

$A$، علاوه بر مربعی بودن، متقارن نیز می باشد ( $A=A^T$ ).

علت این امر این است که همیشه ماتریس متقارنی مانند $B$ وجود دارد به گونه‌ای

که $x^TAx = x^TBx$. برای فهم دلیل این امر اثبات زیر را در نظر بگیرید:

$$x^TAx = (x^TAx)^T$$ → زیرا ترنسپوز یک عدد حقیقی با خودش برابر است.

$$= x^TA^Tx$$

از آنجایی که $x^TAx = x^TA^Tx$، بنابراین میانگین این دو مقدار، خود مقدار $x^TAx$ را می دهد:

$$x^TAx = \frac{x^TAx + x^TA^Tx}{2}$$

$$= x^T \underbrace{\left(\frac{1}{2}A + \frac{1}{2}A^T\right)}_{B} x$$

می دانیم که $A+A^T$ متقارن است، بنابراین $\frac{1}{2}(A+A^T)$ نیز متقارن است.

# Definiteness

با استفاده از quadratic form، می‌توان تعاریف زیر را برای یک ماتریس

مربعی و متقارن $A \in \mathbb{R}^{d \times d}$ و هر بردار غیر صفر $x \in \mathbb{R}^d$ داشت:

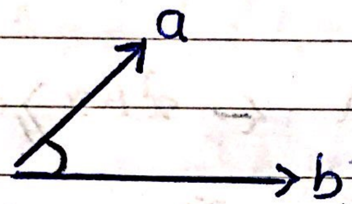
- اگر $x^T A x > 0 , \forall x \neq 0$ ← $A$ یک ماتریس Positive Definite (PD) است.
- اگر $x^T A x \geq 0 , \forall x \neq 0$ ← $A$ یک ماتریس Positive Semidefinite (PSD) است
- اگر $x^T A x < 0 , \forall x \neq 0$ ← $A$ یک ماتریس Negative Definite (ND) است
- اگر $x^T A x \leq 0 , \forall x \neq 0$ ← $A$ یک ماتریس Negative Semidefinite (NSD) است

- در نهایت اگر ماتریسی در هیچ یک از دسته‌های بالا قرار نگیرد، یعنی دو بردار غیرصفر $x_1 , x_2 \in \mathbb{R}^d$ وجود داشته باشد به طوری که $x_1^T A x_1 > 0$ و $x_2^T A x_2 < 0$، ماتریس $A$ یک ماتریس Indefinite است.

نکته- اگر $A$، positive definite باشد، آنگاه $-A$، negative definite است. به طور مشابه اگر $A$، psd باشد، $-A$، nsd است. اگر $A$، indefinite باشد، $-A$ نیز indefinite است.

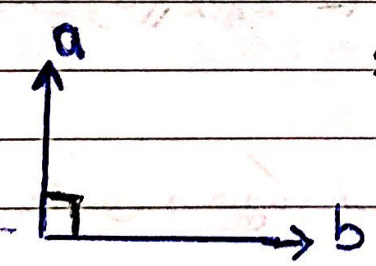
می‌توان quadratic form و definiteness را از نظر هندسی نیز بررسی کرد. می‌دانیم که حاصل ضرب داخلی دو بردار $a, b \in \mathbb{R}^d$، با توجه به زاویه بین دو بردار در یکی از ۳ دسته زیر قرار می‌گیرد:

۱) در صورتیکه زاویه بین a و b کوچکتر از $90^\circ$ باشد، حاصل $a^Tb$ **مثبت** خواهد بود:



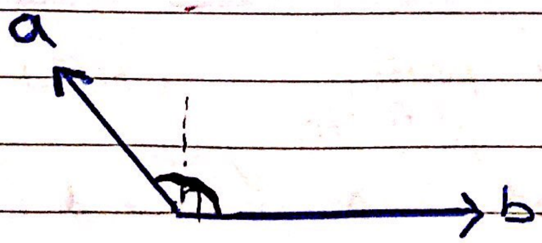

$$a^Tb>0$$

۲) اگر زاویه بین a و b، $90^\circ$ باشد (دو بردار orthogonal هستند)، حاصل $a^Tb$ برابر با **صفر** است:




$$a^Tb=0$$

۳) اگر زاویه بین a و b بزرگتر از $90^\circ$ باشد، حاصل $a^Tb$ **منفی** است:

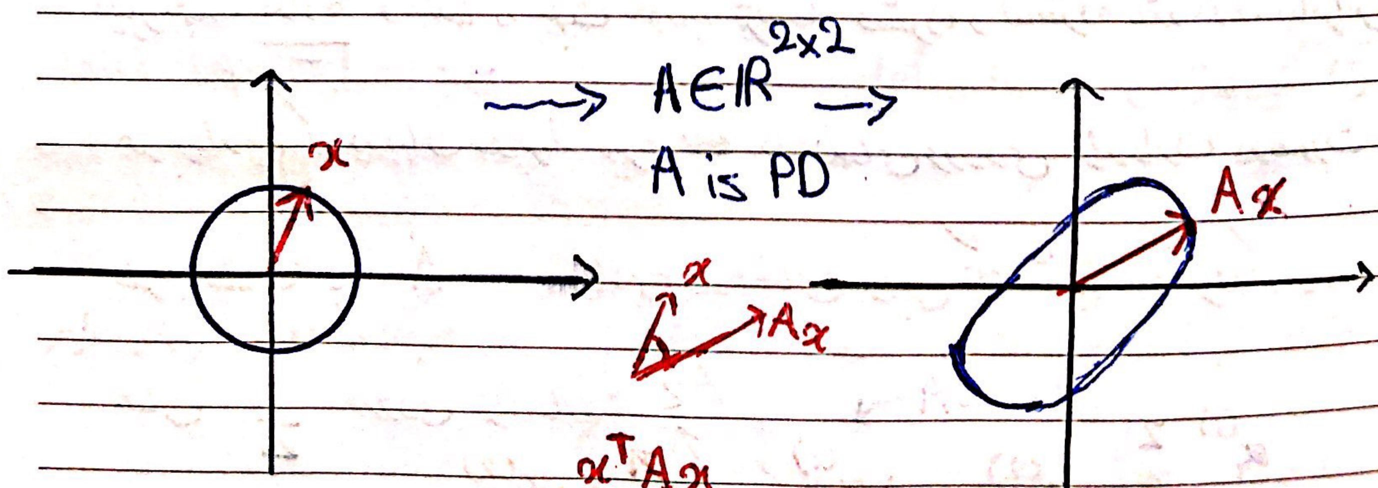



$$a^Tb<0$$

به همین ترتیب می توان quadratic form را به چشم ضرب داخلی میان ورودی ماتریس A و خروجی آن دید : $\underbrace{(x)^T}_{\text{ورودی A}}\underbrace{(Ax)}_{\text{خروجی}}$

بنابراین definiteness یک ماتریس در واقع نشان می دهد که زاویه بین بردارهای ورودی و بردارهای خروجی متناظر پس از ضرب با A ، کوچکتر، بزرگتر و یا برابر با $90^\circ$ است.

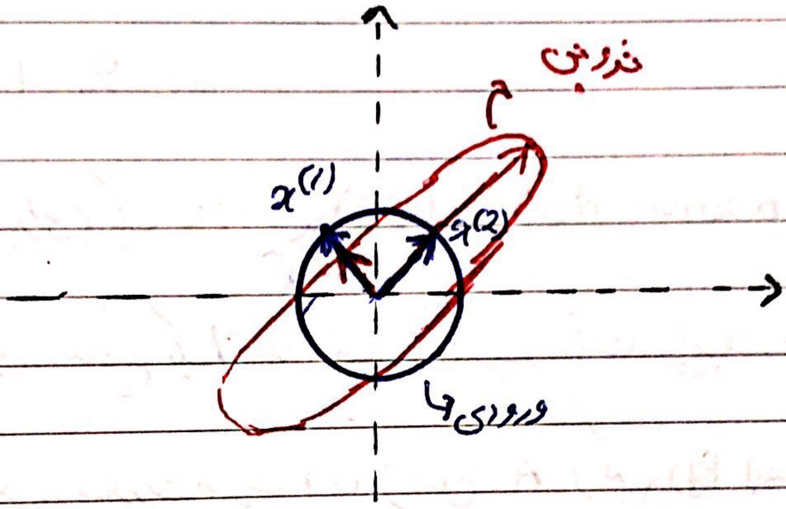
به عنوان مثال ، برای یک ماتریس positive definite (PD) ، تمامی بردارهایی که بر روی دایره‌ای به شعاع یک از مبدأ فضای ورودی آن قرار دارند را در نظر بگیرید. تمامی این بردارها در صورت ضرب با ماتریس A (که PD است) ، برداری را حاصل می شوند که زاویه بین آن و بردار ورودی متناظر کمتر از $90^\circ$ است:



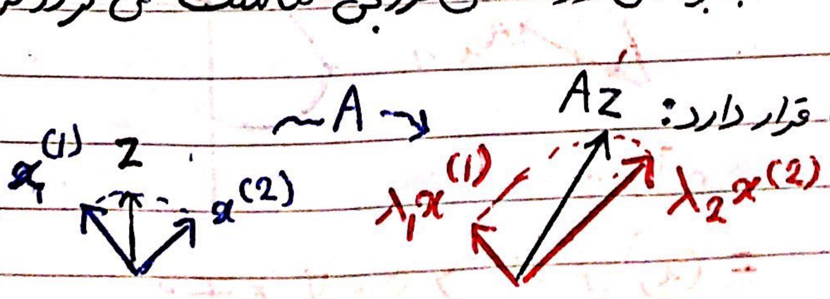

$$x^T A x = (x)^T (Ax) > 0 , \forall x \neq 0$$

بنابراین زاویه بین x و Ax برای همه x های غیر صفر کمتر از $90^\circ$ است

با استفاده از این دیدگاه می‌توان ارتباط کلیدی را میان definiteness و مجموعه مقادیر ویژه ( Spectrum ) یک ماتریس بدست آورد. ماتریس $A\in\mathbb{R}^{2\times2}$ را در نظر بگیرید که تمامی مقادیر ویژه آن مثبت باشد. در این صورت با قرار گرفتن فضای ورودی و خروجی آن روی یکدیگر، شکل زیر حاصل می‌گردد




می‌توان دید که به دلیل مثبت بودن مقادیر ویژه، بردارهای ویژه $x^{(1)}$ و $x^{(2)}$ تغییر جهت نداده و تنها در جهت خود به ترتیب فشرده و گسترده شده‌اند. بنابراین هر برداری که در میان دو بردار ویژه در فضای ورودی قرار دارد، در صورت ضرب شدن با $A$، به برداری در فضای خروجی نگاشت می‌گردد که در همان $\frac{1}{4}$ ورودی قرار دارد:



بنابراین زاویه هر برداری مانند $x$ در فضای ورودی، و بردار خروجی $Ax$،
کمتر از $90^\circ$ خواهد بود. به عبارت دیگر $(x)^T(Ax) > 0, \forall x \neq 0$

پس ماتریس positive definite (PD) است. بدین ترتیب با استفاده
از spectrum ماتریس، definiteness آن مشخص گردید.

به همین ترتیب می‌توان نشان داد:

1) یک ماتریس PD است اگر All Eigenvalues $> 0$

2) یک ماتریس PSD است اگر All Eigenvalues $\geq 0$

3) یک ماتریس ND است اگر All Eigenvalues $< 0$

4) یک ماتریس NSD است اگر All Eigenvalues $\leq 0$

5) یک ماتریس indefinite است اگر some Eigenvalues $> 0$ and some Eigenvalues $< 0$

نکته - از آنجایی که ماتریس های PD و ND هیچگاه مقدار ویژه‌ای برابر با صفر نخواهند داشت، همیشه full rank می‌باشند.

# Decomposition ( تجزیه )

با داشتن یک ماتریس، راه های زیادی برای تجزیه آن وجود دارد. در اینجا دو نوع تجزیه که برای ما اهمیت دارند را بررسی می کنیم:

1) Singular value decomposition (SVD)

2) Eigenvalue decomposition (EVD)

در جدول زیر خلاصه‌ی این تجزیه ها را می توان دید:

| | نوع ماتریس A | تجزیه (Decomposition) |
|---|---|---|
| 1) SVD | هر نوع ماتریس | $A = USV^T$ |
| 2) EVD | فقط ماتریس های مربعی (square) | $A = UDU^T$ |

در جدول بالا ماتریس هایی که با حروف $U$ و $V$ نمایش داده شده اند، ماتریس orthogonal هستند. بنابراین $U^T = U^{-1}$ و $V^T = V^{-1}$.

Scanned with CamScanner

همچنین ماتریس های $S$ و $D$ ماتریس های diagonal هستند.

در تجزیه EVD، ستون های ماتریس $U$، شامل $n$ بردار ویژه مستقل خطی $A$ بوده: $\{v^{(1)}, \dots, v^{(n)}\}$ و مقادیر ویژه متناظر با این بردارهای ویژه، در قطر ماتریس قطری $D$ قرار دارند:

$$U \in \mathbb{R}^{n\times n}, \quad U = \begin{bmatrix} | & | & & | \\ v^{(1)} & v^{(2)} & \cdots & v^{(n)} \\ | & | & & | \end{bmatrix}$$

$$D \in \mathbb{R}^{n\times n}, \quad D = \begin{bmatrix} \lambda^{(1)} & & 0 \\ & \lambda^{(2)} & \\ & & \ddots \\ 0 & & & \lambda^{(n)} \end{bmatrix}$$

در تجزیه SVD، $U$ و $V$ می توانند هر ماتریس orthagonal بوده و به دلیل اینکه $A$ می تواند ماتریس غیرمربعی باشد، ابعاد $U$ و $V$ می توانند متفاوت باشند.

به عنوان مثال ممکن است: $U \in \mathbb{R}^{m\times m}$ و $S \in \mathbb{R}^{m\times n}$ و $V \in \mathbb{R}^{n\times n}$

برای ماتریس $A \in \mathbb{R}^{m\times n}$

دقت کنید که ماتریس قطری $S$، می تواند مربعی نباشد. مؤلفه هایی که در قطر اصلی ماتریس $S$ قرار دارند singular values ماتریس $A$ نامیده می شوند. همچنین ستون های ماتریس $U$، left-singular vectors و ستون های

ماتریس V، right-singular vectors نامیده می شوند.

نکته - در SVD، همیشه مؤلفه‌های قطر اصلی ماتریس S که همان singular values هستند، حقیقی می باشند (برای هر ماتریس A). این درحالی است که مقادیر ویژه ممکن است اعداد مختلط باشند (برای ماتریس های مربعی غیرمتقارن)

برای درک بهتر مفهوم تجزیه یک ماتریس، می توانیم دوباره به تفسیر هندسی و حالتی که ماتریس را به چشم یک تابع می دیدیم برگردیم. دیدیم که ماتریس A را می توان به چشم تابعی دید که با دریافت بردار $x$ به عنوان ورودی، عملیاتی انجام داده و بردار $y=A(x)$ را به عنوان خروجی می دهد. زمانی که ماتریس A تجزیه می گردد، درواقع به این معنا است که عملیاتی که ماتریس A بر روی بردار $x$ انجام می دهد، قابل تجزیه به سه زیرعملیات است. در جدول زیر این سه زیرعملیات به تفکیک نشان داده شده اند:

| | step1 | step2 | step 3 |
|---|---|---|---|
| SVD | Rotation and Reflection by $V^T$ | scaling along axes using diagonal matrix S | Rotation and Reflection by U |
| EVD | Rotation and Reflection by $U^T$ | scaling along axes using diagonal matrix D | Rotation and Reflection inverse of $U^{-1}$ by U |

همانطور که در جدول قابل مشاهده است، در زیرعملیات اول (step 1)، بردار ورودی با یک ماتریس orthogonal ضرب می شود. ضرب با ماتریس orthogonal معادل انجام عملیات Rotation and Reflection (چرخش و بازتاب) بر روی بردار ورودی است.

علت این امر این است که ماتریس های orthogonal اندازه و زوایای میان بردارها را در فضا تغییر نداده و تنها چرخش و بازتاب ایجاد می کنند.

برای درک این امر دو بردار فرضی $x, y \in \mathbb{R}^n$ را در نظر بگیرید. حاصل ضرب داخلی این دو بردار ($x^T y$) به زاویه میان این دو و اندازه دو بردار وابسته است. حال فرض کنید دو بردار با ماتریس orthogonal مانند $U$ ضرب شوند. در اینصورت برای ضرب داخلی دو بردار حاصل داریم:

$$(Ux)^T (Uy) = x^T U^T U y$$

می دانیم که برای ماتریس های orthogonal: $U^T U = I = U U^T$، بنابراین $U^T = U^{-1}$. پس داریم:

$$x^T \underbrace{U^T U}_{I} y = x^T y$$

بنابراین حاصل ضرب داخلی دو بردار، با ضرب در ماتریس $U$ ثابت مانده که

نشان دهنده‌ی حفظ شدن زاویه بین دو بردار و اندازه آنها است.

- در زیر عملیات دوم، بردار حاصل از زیر عملیات اول با یک ماتریس قطری ($S$ و $D$) ضرب می گردد. این عملیات معادل تغییر مقیاس هر یک از ابعاد بردار می باشد. به عبارت دیگر با ضرب بردار $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ با ماتریس قطری $D = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{bmatrix}$، $x_1$، برابر $d_1$، $x_2$، برابر $d_2$ و در نهایت $x_n$، برابر $d_n$ برابر می گردد.

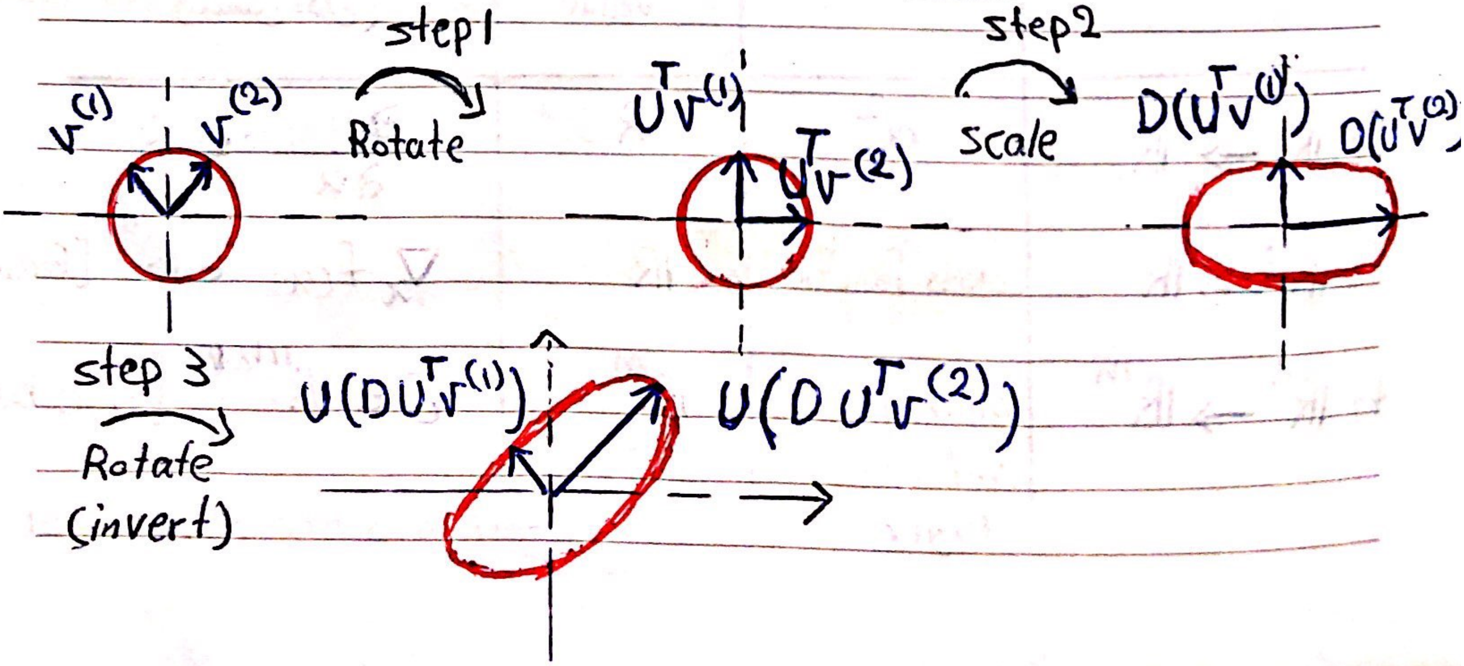
نکته- زیر عملیات دوم برای SVD، همیشه معادل عملیات scaling است، زیرا مقادیر singular values که در قطر اصلی ماتریس $S$ قرار می گیرند همیشه اعداد حقیقی اند. این در حالی است که در EVD، در صورت مختلط بودن مقادیر ویژه که در قطر اصلی $D$ قرار می گیرند، علاوه بر تغییر مقیاس (scaling) می تواند مقداری چرخش (Rotation) نیز در صورت ضرب با اعداد مختلط حاصل گردد.

- نهایتاً با ضرب بردار حاصل از زیر عملیات دوم با ماتریس orthogonal

دیگری ( V و U ) ، در زیرعملیات سوم دوباره Rotation and Reflection داریم. دقت کنید که در SVD ، چرخش مرحله اول و سوم با یکدیگر ارتباطی ندارند اما در EVD چرخش زیرعملیات سوم دقیقاً معکوس چرخش زیرعملیات اول است.

**نکته-** برای ماتریس های مربعی و متقارن ، نتیجه حاصل از SVD و EVD یکسان است. به عبارت دیگر سه ماتریس حاصل در هر دو تجزیه مشابه اند. همچنین دقت کنید که برای ماتریس های غیر مربعی ، تنها SVD قابل انجام است.

در شکل زیر می توانید درک هندسی از سه عملیات را در تجزیه EVD در ساده ترین حالت یعنی برای ماتریس مربعی و متقارن A ببینید ( $A\in\mathbb{R}^{2\times2}$ و $A=A^T$ ):



# Matrix Calculus

می دانیم که زمانیکه می نویسیم $f: \mathbb{R} \to \mathbb{R}$ ، به این معنا است که تابعی به نام f وجود دارد و به عنوان ورودی عددی حقیقی دریافت کرده و به عنوان خروجی عددی حقیقی می دهد. اما یک تابع می تواند به عنوان ورودی بردار و یا ماتریس دریافت کرده و به عنوان خروجی بردار یا ماتریس برگرداند. در اینصورت مشتق چنین تابعی چگونه محاسبه شده و چه نامیده می شود؟

در جدول زیر خلاصه ای از انواع مختلف تابع بر اساس نوع ورودی و خروجی نشان داده شده است:

| Function | Example | Output Value | 1st derivative (مشتق اول) |
|---|---|---|---|
| $f: \mathbb{R} \to \mathbb{R}$ | $x^2$ | $\mathbb{R}$ | $\frac{\partial f(x)}{\partial x} \in \mathbb{R}$ |
| $f: \mathbb{R}^n \to \mathbb{R}$ | loss function | $\mathbb{R}$ | $\nabla_x f(x) \in \mathbb{R}^n$ [Gradient] |
| $f: \mathbb{R}^n \to \mathbb{R}^m$ | Neural Network layer | $\mathbb{R}^m$ | $J \in \mathbb{R}^{m \times n}$ [Jacobian] |

ادامه ی جدول صفحه ی بعد است

| Function | 2nd derivative (مشتق دوم) |
| --- | --- |
| $f: \mathbb{R} \to \mathbb{R}$ | $\dfrac{\partial^2 f(x)}{\partial x^2} \in \mathbb{R}$ |
| $f: \mathbb{R}^n \to \mathbb{R}$ | $H = \nabla_x^2 f(x) \in \mathbb{R}^{n \times n}$ [Hessian] |
| $f: \mathbb{R}^n \to \mathbb{R}^m$ | در این مرور بررسی نمی‌شود |

توضیح جدول و نکات مهم:

با داشتن تابعی مانند $f: \mathbb{R}^n \to \mathbb{R}$ ، که برداری مانند $x \in \mathbb{R}^n$ را به عنوان ورودی گرفته و عددی حقیقی را در خروجی ما می‌دهد، می‌توان ماتریس به نام Gradient محاسبه کرد که برابر است با مشتق جزئی تابع f نسبت به هر یک از n ورودی بردار x . Gradient تابع $f(x)$ نسبت به بردار x را با نماد $\nabla_x f(x)$ نمایش می‌دهیم:

$$\nabla_x f(x) \in \mathbb{R}^n = \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

نکته - می توان Gradient تابعی که به عنوان ورودی ماتریس می گیرد و

به عنوان خروجی عددی حقیقی می دهد را نیز با توجه به تعریف صفحه قبل محاسبه

نمود. به عنوان مثال برای $f: \mathbb{R}^{m \times n} \to \mathbb{R}$ داریم:

(ورودی تابع یک ماتریس مانند $A \in \mathbb{R}^{m \times n}$ است)

$$\nabla_A f(A) \in \mathbb{R}^{m \times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial a_{11}} & \cdots & \frac{\partial f(A)}{\partial a_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial a_{m1}} & \cdots & \frac{\partial f(A)}{\partial a_{mn}} \end{bmatrix}$$

نکته - بردار گرادیان حاصل برای ورودی $x \in \mathbb{R}^n$، در واقع جهتی را می دهد

که حرکت در آن جهت باعث بیشترین افزایش تابع $f(x)$ در نقطه $x$ می شود.

همچنین برای تابع $f: \mathbb{R}^n \to \mathbb{R}$، ماتریس Hessian شامل

تمامی مشتق های جزئی دوم تابع $f(x)$ نسبت به تمامی مؤلفه های $x$

می باشد. Hessian را به صورت $H = \nabla_x^2 f(x)$

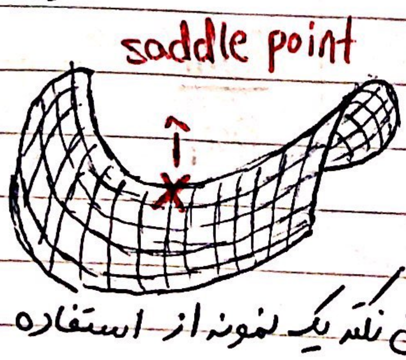
نمایش می دهیم:

$$H = \nabla_x^2 f(x) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2}, & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2}, & \cdots, & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1}, & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

دقت کنید که ماتریس $H$ یک ماتریس مربعی و متقارن است. ماتریس‌های مربعی و متقارن را با نماد $\mathbb{S}^n$ نیز نشان می‌دهند یعنی: $H \in \mathbb{S}^n$

نکته- اگر Hessian یک تابع positive definite (PD) / positive semidefinite (PSD) باشد، تابع convex (محدب) بوده و اگر negative semidefinite (NSD) / negative definite (ND) باشد، تابع concave (مقعر) است.

اگر تابعی indefinite باشد، تابع دارای saddle points یا نقاط زینی است:




یعنی تابع در راستای یک بعد محدب و در راستای بعدی دیگر مقعر است. این نکته یک نمونه از استفاده ما از جبر خطی

در یادگیری ماشین است. با آنالیز کردن Hessian یک تابع می‌توان فهمید که آیا دارای ماکسیمم یا مینیمم تابع است یا ~~...~~ خیر. این امر در بهینه‌سازی تابع اهمیت ویژه‌ای دارد.

---

در نهایت Jacobian یک تابع $f: \mathbb{R}^n \to \mathbb{R}^m$، ماتریسی است شامل مشتق جزئی تمامی مؤلفه‌های بردار خروجی ($\mathbb{R}^m$) نسبت به تمامی مؤلفه‌های بردار ورودی ($\mathbb{R}^n$). Jacobian را می‌توان برحسب Gradian تصور کرد. اگر گرادیان هر یک از مؤلفه‌های خروجی را نسبت به بردار ورودی محاسبه کرده و به‌صورت سطری زیر هم قرار دهیم، ماتریس Jacobian حاصل می‌گردد:

$$J = \begin{bmatrix} \nabla_x^T f_1(x) \\ \vdots \\ \nabla_x^T f_m(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

نکته ـ چند مثال مهم :

- $$\nabla_x b^T x = \begin{bmatrix} \vdots \\ \frac{\partial}{\partial x_i}(b^T x) \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \frac{\partial}{\partial x_i}(b_1 x_1 + \cdots b_i x_i + \cdots b_n x_n) \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \vdots \\ b_i \\ \vdots \end{bmatrix} = b$$

- $$\nabla_A \log|A| = A^{-1}$$

- Product Rule

می دانیم که برای مشتق داریم :

$$\frac{d\, f(x)\, g(x)}{dx} = \left(\frac{df(x)}{dx}\right) g(x) + f(x)\left(\frac{d\, g(x)}{dx}\right)$$

به همین صورت برای gradient داریم :

$$\nabla_x \underbrace{x^T}_{f(x)} \underbrace{Ax}_{g(x)} = \nabla_x x^T A x + \nabla_x x^T A x$$

$$= Ax + A^T x = (A + A^T) x$$

(Symmetric) $A = A^T$ اگر

$$= 2Ax$$