به نام خدا

دانشگاه صنعتی شریف - دانشکده مهندسی کامپیوتر

مقدمه‌ای بر یادگیری ماشین

فصل اول: مروری بر ریاضیات یادگیری ماشین

جبرخطی

نویسندگان: کوروش مسلمی

1.جبرخطی

در اینجا به یادآوری کوتاهی از جبرخطی در راستای اهداف درس می پردازیم

1.1.بردارها و فضای برداری

به یک n تایی به صورت $X = [x_1, x_2, \cdots, x_n]$ یک **بردار سطری** گفته می شود و می گوییم X یک بردار $1 \times n$ است.

مشابها یک n تایی به صورت $Y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ **بردار ستونی** $n \times 1$ است.

در ادامه به عنوان مثال بردارهای ستونی و سطری تولید می کنیم و شیوه دسترسی به المان آن‌ها را نشان می دهیم.

```
import numpy as np
X = np.arange(3)
Y = np.empty([3,1])
print(X,X.shape)
print(Y,Y.shape)
print(X[1],Y[1])
```

```
[0 1 2] (3,)
[[0.e+000]
 [5.e-324]
 [1.e-323]] (3, 1)
1 [5.e-324]
```

یک **فضای برداری** مجموعه ای مانند S است که عناصر آن بردار هستند. اعمال جمع و ضرب اسکالری برای این عناصر تعریف می شود که در شرایط ذیل صدق می کنند.(U ، V و W عناصر دلخواه S ، c و d اسکالر هستند)

1. حاصل جمع U + V وجود دارد و متعلق به S است.(S تحت جمع بسته است)
2. cU عنصری از S است.(S تحت ضرب اسکالر بسته است.)
3. $U + V = V + U$ (ویژگی جابجایی)
4. $U + (V + W) = (U + V) + W$ (ویژگی شرکت پذیری)
5. عنصری از V، که بردار صفر نامیده می شود، با 0 نشان داده می شود، وجود دارد به قسمی که $U + 0 = U$
6. متناظر هر عنصر U از S عنصری وجود دارد که قرینه U نامیده و با U− نشان داده می شود به قسمی که:
$$U + (-U) = 0$$
7. $\mathrm{c}(U + V) = \mathrm{c}U + \mathrm{c}V$
8. $(\mathrm{c} + \mathrm{d})U = \mathrm{c}U + \mathrm{d}U$
9. $\mathrm{c}(\mathrm{d}U) = (\mathrm{cd})U$
10. $۱U = U$

فضاهای برداری می توانند شامل فضاهای برداری دیگری نیز باشند. اگر V یک فضای برداری باشد آنگاه $S \subseteq V$ یک **زیرفضای** V نامیده می شود اگر:

1. $0 \in S$
2. S تحت جمع بسته باشد: $x, y \in S \Rightarrow x + y \in S$
3. S تحت ضرب اسکالر بسته باشد: $x \in S, \alpha \in \mathbf{R} \Rightarrow \alpha S \in S$

به عنوان مثال یک خط که از مبدا می گذرد زیر فضای R^2 است.

فرض کنیم $V_1, V_2, \cdots, V_m$ بردارهایی در فضای برداری S باشند. گوییم V، برداری در S، یک **ترکیب خطی** از $V_1, V_2, \cdots, V_m$ است هرگاه اسکالرهایی چون $c_1, c_2, \cdots, c_m$ موجود باشند به قسمی که V را بتوان چنین نوشت:

$$V = c_1 V_1 + c_2 V_2 + \cdots + c_m V_m$$

مجموعه بردارهای $V_1, V_2, \cdots, V_m$ از یک فضای برداری را وقتی یک **مولد** این فضا می نامیم که هر بردار فضا ترکیبی از این بردارها باشد. یک مجموعه مولد از بردارها به یک معنی فضای برداری مربوطه را معرفی می کند زیر هر بردار آن فضا را می توان از این بردار ها بدست آورد.

مجموعه بردارهای $V_1, V_2, \cdots, V_n$ را **مستقل خطی** گوییم هرگاه از معادله

$$c_1 V_1 + c_2 V_2 + \cdots + c_n V_n = 0 \qquad (1)$$

نتیجه شود که

$$c_1 = c_2 = \cdots = c_n = 0$$

چنانچه حداقل یک $c_i \neq 0$ وجود داشته باشد که معادله (1) برقرار باشد گوییم مجموعه بردارهای $V_1, V_2, \cdots, V_n$ وابسته خطی هستند.

مجموعه ای متناهی از بردارها مانند $\{V_1, \cdots, V_m\}$ را یک **پایه** فضای برداری S نامیم هرگاه این مجموعه مولد S و مستقل خطی باشد.

مثال) آیا بردارهای $\{(1,-1,1),(-1,-1,1),(0,1,1)\}$ برای $\mathbf{R}^3$ پایه هستند؟

حل)

اسکالرهای c_1, c_2, c_3 باید وجود داشته باشند که

$$c_1(1,-1,1)+c_2(-1,-1,1)+c_3(0,1,1)=(x_1,x_2,x_3)$$
$$\Rightarrow \begin{cases} c_1-c_2=x_1 \\ -c_1-c_2+c_3=x_2 \\ c_1+c_2+c_3=x_3 \end{cases} \Rightarrow \begin{cases} c_1=\frac{2x_1-x_2+x_3}{4} \\ c_2=\frac{-2x_1-x_2+x_3}{4} \\ c_3=\frac{x_2+x_3}{2} \end{cases}$$

اکنون مستقل خطی بودن را بررسی می کنیم

$$b_1(1,-1,1)+b_2(-1,-1,1)+b_3(0,1,1)=0$$
$$\Rightarrow \begin{cases} b_1-b_2=0 \\ -b_1-b_2+b_3=0 \\ b_1+b_2+b_3=0 \end{cases} \quad \Rightarrow b_1=b_2=b_3=0$$

پس این مجموعه بردارها پایه $\mathbf{R}^3$ هستند.

2.1.ماتریس‌ها

یک ماتریس مجموعه ای از بردارها است مثلا اگر

$$A=\begin{bmatrix} 1 & 2 & 9 \\ 0 & -1 & 6 \\ 1 & 2 & 7 \end{bmatrix}_{3\times 3}, \qquad B=\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}_{2\times 2}$$
$$C=\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}_{2\times 3}, \qquad D=\begin{bmatrix} 1 & 9 \\ -1 & 0 \\ 2 & 6 \end{bmatrix}_{3\times 2}$$

آنگاه:

A ماتریسی است که 3 سطر و 3 ستون دارد

B ماتریسی است که 2 سطر و 2 ستون دارد

C ماتریسی است که 2 سطر و 3 ستون دارد

D ماتریسی است که 3 سطر و 2 ستون دارد

در ادامه یک نمونه ماتریس تولید می کنیم.

```
A = np.arange(6).reshape(3, 2)
A
```

```
array([[0, 1],
       [2, 3],
       [4, 5]])
```

ترانهاده ماتریس A را با A^T نمایش می دهیم که از جابجایی سطرها و ستون‌های آن ایجاد می شود پس اگر

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

آنگاه

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

در ادامه برابری ترانهاده یک ماتریس متقارن با خودش را بررسی می کنیم

```
A = np.array([[1, 2, 3], [2, 0, 4], [3, 4, 5]])
A == A.T
```

```
array([[ True,  True,  True],
       [ True,  True,  True],
       [ True,  True,  True]])
```

فرض کنیم A یک ماتریس $m \times n$ باشد، سطرهای A را می توانیم به مثابه بردار های سطری $r_1, r_2, \cdots, r_m$ و ستون‌های آن را به مثابه بردارهای ستونی $c_1, c_2, \cdots, c_n$ بینگاریم. هر بردار سطری شامل n مولفه و هر بردار ستونی شامل m مولفه می باشد. بردارهای سطری زیرفضایی از $\mathbf{R}^n$ را گسترش می دهند که فضای سطری A نامیده می شود.

مثال) ماتریس زیر را در نظر می گیریم

$$A = \begin{bmatrix} 1 & 4 \\ 5 & 9 \end{bmatrix}$$

بردارهای سطری A عبارتند از

$$r_1 = (1, 4), \qquad r_2 = (5, 9)$$

این بردارها زیرفضایی از $\mathbf{R}^2$ تولید می کنند که **فضای سطری** A نامیده می شود. از طرفی بردارهای ستونی A عبارتند از

$$c_1 = \begin{bmatrix} 1 \\ 5 \end{bmatrix}, \qquad c_2 = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$$

این بردارها زیرفضایی از $\mathbf{R}^2$ تولید می کنند که **فضای ستونی** A نامیده می شود.

قضیه: فضای سطری و ستونی ماتریس A دارای یک بعد هستند.

قضیه: بعد فضای سطری(ستونی) ماتریس A را **رتبه** A می نامیم. رتبه A را با نماد $Rank(A)$ نشان می دهیم.

مثال) رتبه ماتریس زیر را پیدا کنید.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 2 & 5 & 8 \end{bmatrix}$$

حل) با بررسی در می یابیم که سطر سوم این ماتریس ترکیب خطی از دو سطر آن است:

$$(2,5,8) = 2(1,2,3) + (0,1,2)$$

در این صورت سه سطر این ماتریس وابسته ی خطی هستند. در نتیجه رتبه این ماتریس کوچک تر از 3 است. چون (1,2,3) مضرب اسکالری از (0,1,2) نیست، این دو بردار مستقل خطی هستند. این دو بردار تشکیل یک پایه برای فضای سطری A می دهند. از این رو $Rank(A) = 2$

برای هر ماتریس $m \times n$ دو زیر فضای مهم به صورت زیر وجود دارند

$$R(A) = \{B \in \mathbf{R}^m \mid B = AX, X \in \mathbf{R}^n\} \qquad \text{فضای برد A}$$
$$N(A) = \{X \in \mathbf{R}^n \mid AX = 0\} \qquad \text{فضای پوچ یا کرنل A}$$

توجه کنید بعد $N(A)$ را با $dim(N(A))$ نشان می دهیم.

قضیه: فرض کنید A ماتریس $m \times n$ باشد آنگاه

$$Rank(A) + dim(N(A)) = n$$

تذکر: رتبه A همان بعد فضای برد A می باشد.

مثال: فضای برد و پوچ ماتریس زیر را محاسبه کنید.

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

برای محاسبه فضای پوچ A قرار می دهیم:

$$AX = 0 \Rightarrow \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} x_1 + 2x_2 - x_3 = 0 \\ 2x_1 + 4x_2 - 2x_3 = 0 \end{cases} \longrightarrow x_1 + 2x_2 - x_3 = 0$$

اگر $x_3 = \alpha$ و $x_2 = \beta$ آنگاه $x_1 = -2\beta + \alpha$ پس

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2\beta + \alpha \\ \beta \\ \alpha \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

بنابراین $N(A)$ شامل تمام بردار هایی به شکل فوق است که $\alpha, \beta \in \mathbf{R}$ پس

$$N(A) = \left\{ \alpha \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \mid \alpha, \beta \in \mathbf{R} \right\}$$

لذا مجموعه $\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}$ پایه ای برای این فضا و در نتیجه $dim(N(A)) = 2$ از اینرو قضیه قبل نتیجه می دهد

$$Rank(A) = n - dim(N(A)) = 3 - 2 = 1$$

از طرفی برای فضای برد داریم:

$$b = AX = \begin{bmatrix} x_1 + 2x_2 - x_3 \\ 2x_1 + 4x_2 - 2x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2x_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - x_3 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$
$$= (x_1 + 2x_2 - x_3) \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \gamma \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \gamma \in \mathbf{R}$$

بنابراین

$$R(A) = \left\{ \gamma \begin{bmatrix} 1 \\ 2 \end{bmatrix} \;\mid\; \gamma \in \mathbf{R} \right\}$$

فضای برد A است. واضح است که $dim(R(A)) = 1$ لذا از این هم می توان متوجه شد که

$$Rank(A) = dim(R(A)) = 1$$

دترمینان ماتریس 2×2، A را با $det(A)$ نشان داده و به صورت زیر تعریف می کنیم:

$$det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}$$

کهاد درایه a_{ij} را با M_{ij} نشان می دهیم و عبارت است از دترمینان ماتریس حاصل از حذف ماتریس ردیف i ام و ستون j ام ماتریس A .

همسازه درایه a_{ij} را با C_{ij} نشان می دهیم و به صورت زیر تعریف می کنیم:

$$C_{ij} = (-1)^{i+j} M_{ij}$$

توجه کنید که کهاد و همسازه یک درایه حداکثر تفاوتشان در یک علامت است.

$$C_{ij} = +M_{ij} \text{ یا } - M_{ij}$$

مثال) کهاد و همسازه درایه های a_{12}, a_{21} را برای ماتریس زیر بدست آورید

$$A = \begin{bmatrix} 2 & -1 & 5 \\ 0 & 3 & 4 \\ 2 & 2 & -2 \end{bmatrix}$$

حل: با بکارگیری تعریف اخیر داریم

$$\begin{aligned} M_{12} &= det\left(\begin{bmatrix} 0 & 4 \\ 2 & -2 \end{bmatrix}\right) = 0 \times (-2) - (4 \times 2) = -8 \\ C_{12} &= (-1)^{1+2} M_{12} = -1 \times -8 = 8 \\ M_{33} &= det\left(\begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix}\right) = 2 \times 3 - (-1) \times 0 = 6 \\ C_{33} &= (-1)^{3+3} M_{33} = 1 \times 6 = 6 \end{aligned}$$

دترمینان ماتریس مربعی A برابر است با مجموع حاصل ضرب درایه های هر سطر یا هر ستون در همسازه مربوط به آن درایه. بنابراین دترمینان ماتریس A برحسب بسط سطر i ام برابر است با

$$det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

مثال) مطلوب است محاسبه دترمینان ماتریس A برحسب

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 13 & -2 & 5 \\ -1 & 1 & 0 \end{bmatrix}$$

سطر سوم:

$$(-1)^4 \times (-1) \times det\left(\begin{bmatrix} 0 & 3 \\ -2 & 5 \end{bmatrix}\right) + (-1)^5 \times 1 \times det\left(\begin{bmatrix} 1 & 3 \\ 13 & 5 \end{bmatrix}\right) = 28$$

سطر دوم:

$$(-1)^3 \times 13 \times det\left(\begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}\right) + (-1)^4 \times -2 \times det\left(\begin{bmatrix} 1 & 3 \\ -1 & 0 \end{bmatrix}\right) + (-1)^5 \times 5 \times det\left(\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\right) = 28$$

ستون اول:

$$(-1)^2 \times 1 \times det\left(\begin{bmatrix} -2 & 5 \\ 1 & 0 \end{bmatrix}\right) + (-1)^3 \times 13 \times det\left(\begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}\right) + (-1)^4 \times -1 \times det\left(\begin{bmatrix} 0 & 3 \\ -2 & 5 \end{bmatrix}\right) = 28$$

ستون دوم:

$$(-1)^3 \times 0 + (-1)^4 \times -2 \times det\left(\begin{bmatrix} 1 & 3 \\ -1 & 0 \end{bmatrix}\right) + (-1)^5 \times 1 \times det\left(\begin{bmatrix} 1 & 3 \\ 13 & 5 \end{bmatrix}\right) = 28$$

در ادامه چند خاصیت از دترمینان می بینیم. می توانید برای تمرین آن ها را اثبات کنید.

1. $det(I) = 1$
2. $det(A^T) = det(A)$
3. $det(AB) = det(A)det(B)$
4. $det(A^{-1}) = det(A)^{-1}$
5. $det(\alpha A) = \alpha^n det(A)$

اثر یک ماتریس که با نماد $trace(A)$ نمایش داده می شود، به صورت زیر است

$$trace(A) = \sum_{i=1}^{n} a_{ii}$$

تمرین) خواص زیر را برای اثر ماتریس نشان دهید.

1. trace(A+B) = trace(A) + trace(B)
2. trace(α A) = α trace(A)
3. trace(A^T) = trace(A)
4. trace(ABCD) = trace(BCDA) = trace(CDAB) = trace(DABC)

اگر A یک ماتریس $n \times n$ باشد، آنگاه عبارات زیر معادل هستند:

1. ماتریس A نامنفرد(ناتکین، معکوس پذیر) است
2. دترمینان ماتریس A مخالف صفر است
3. دستگاه $AX = B$ به ازای هر b جواب منحصر به فرد دارد
4. $Rank(A) = n$
5. $AX = 0$ جواب بدیهی صفر دارد
6. ماتریس A با ماتریس I_n هم ارز سطری است
7. ستون‌های ماتریس A مستقل خطی هستند
8. سطرهای ماتریس A مستقل خطی هستند
9. $dim(N(A)) = 0$
10. مقادیر ویژه ماتریس A مخالف صفر هستند
11. یکتا جواب دستگاه $X = A^{-1}B$ است.
12. $dim(R(A)) = n$

ماتریس مربعی A را **مثبت معین** گوییم هرگاه برای هر بردار $X \neq 0$ داشته باشیم :

$$X^T(\frac{A+A^T}{2})X > 0$$

مثال) نشان دهید ماتریس داده شده مثبت معین است. $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

حل) باید نشان دهیم برای هر $0 \neq X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ داریم : $X^T(\frac{A+A^T}{2})X > 0$

ابتدا توجه داشته باشید که

$$\frac{1}{2}(A+A^T) = \frac{1}{2}\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

از طرفی داریم:

$$X^T(\frac{A+A^T}{2})X = [x_1\ x_2]\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [x_1\ x_2]\begin{bmatrix} x_1 + \frac{1}{2}x_2 \\ \frac{1}{2}x_1 + x_2 \end{bmatrix}$$
$$= x_1^2 + x_1x_2 + x_2^2 = \frac{1}{2}(x_1^2 + x_2^2 + (x_1+x_2)^2) > 0$$

لذا ماتریس A مثبت معین است.

توجه) در بیشتر موارد کاربردی برای مثبت معین بودن، ماتریس A متقارن است و لذا تعریف به صورت زیر خلاصه می‌شود.

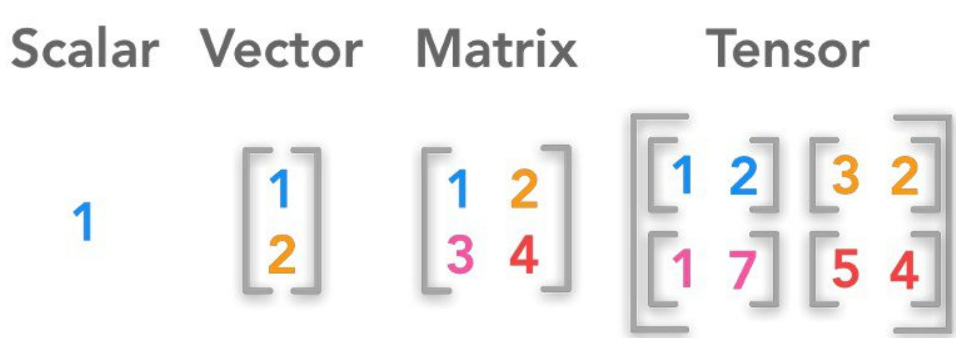
$$A^T = A \Longrightarrow X^T(\frac{A+A^T}{2})X = X^T(\frac{A+A}{2})X = X^TAX > 0$$

بنابراین ماتریس متقارن A را مثبت معین گوییم هرگاه برای هر $X \neq 0$ داشته باشیم :

$$X^TAX > 0$$

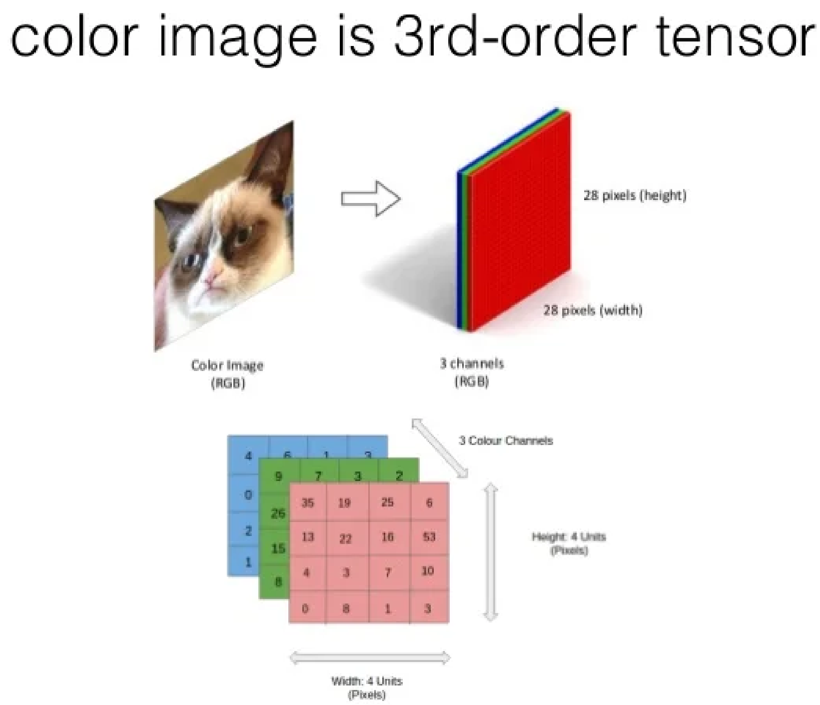
3.1.تنسور ها

تنسور ها تعمیمی از اسکالر ها، بردارها و ماتریس ها برای ابعاد بالاتر از 2 هستند.



فرض کنید عکسی سیاه و سفید (grayscale) در اختیار دارید در این صورت می توانید آن را در یک ماتریس ذخیره کنید. به این صورت که به هر پیکسل مقداری بین 0 تا 255 اختصاص دهید و در یک خانه ماتریس ذخیره کنید. در یادگیری ماشین و یادگیری عمیق عموما با داده‌ها با ابعاد بالا سر و کار داریم. به طور مثال برای ذخیره سازی یک عکس رنگی در

فضای رنگی RGB نیاز به یک تنسور 3 بعدی داریم.همچنین برای نگهداری مجموعه ای از عکس‌ها به تنسور 4 بعدی نیاز



در ادامه یک تنسور سه بعدی تولید می کنیم

```
np.arange(24).reshape(2, 3, 4)
```

```
array([[[ 0,  1,  2,  3],
        [ 4,  5,  6,  7],
        [ 8,  9, 10, 11]],

       [[12, 13, 14, 15],
        [16, 17, 18, 19],
        [20, 21, 22, 23]]])
```

4.1.عملیات روی تنسورها

ابتدا ضرب درایه‌ای (Hadamard product) را معرفی می کنیم. فرض کنید $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ در این صورت داریم:

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \dots & a_{2n}b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \dots & a_{mn}b_{mn} \end{bmatrix}.$$

مثال:

```
A = np.arange(6, dtype=np.float32).reshape(2, 3)
B = A.copy()  # Assign a copy of `A` to `B` by allocating new memory
```

```
A * B
```

```
array([[ 0.,  1.,  4.],
       [ 9., 16., 25.]], dtype=float32)
```

یکی از اساسی ترین عملیات‌ها در جبرخطی ضرب نقطه ای (Dot Product) یا ضرب داخلی (Inner Product) می باشد.
فرض کنید $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ در این صورت تعریف می کنیم:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{d} x_i y_i$$

به عبارت دیگر ضرب داخلی روی فضای برداری V تابعی است به فرم $\mathbf{R} \rightarrow \mathbf{V} \times \mathbf{V}$: $\langle . , . \rangle$ که در شرایط زیر صدق می کند:

1. $\langle X, X \rangle \geq 0$
2. $\langle X, X \rangle = 0 \Leftrightarrow X = 0$
3. $\langle X + Y, Z \rangle = \langle X, Z \rangle + \langle Y, Z \rangle$
4. $\langle \alpha X, Z \rangle = \alpha \langle X, Z \rangle$
5. $\langle X, Y \rangle = \langle Y, X \rangle$

همانطور که در ادامه می بینیم V یک فضای متریک محسوب می شود.

یکی از کاربردهای ضرب نقطه ای محاسبه میانگین وزن دار است. فرض کنید $\mathbf{x} \in \mathbb{R}^n,\ \mathbf{w} \in \mathbb{R}^n,\ \sum_{i=1}^{n} w_i = 1$ که x بردار مقادیر و w بردار وزن‌های نامنفی مقادیر است در این صورت $\mathbf{x}^\top \mathbf{w}$ همان میانگین وزن دار است.

مثال:

```
x = np.arange(3, dtype=np.float32)
w = np.array([0.5,0.2,0.3])
print(np.dot(x, w)) # method 1
print(np.sum(x * w)) # method 2
```

```
0.8
0.8
```

در ادامه به ضرب ماتریس-بردار می پردازیم. فرض کنید A و x به ترتیب ماتریس $m \times n$ و بردار n بعدی هستند. قرار می دهیم:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}^\top_1 \\ \mathbf{a}^\top_2 \\ \vdots \\ \mathbf{a}^\top_m \end{bmatrix},$$

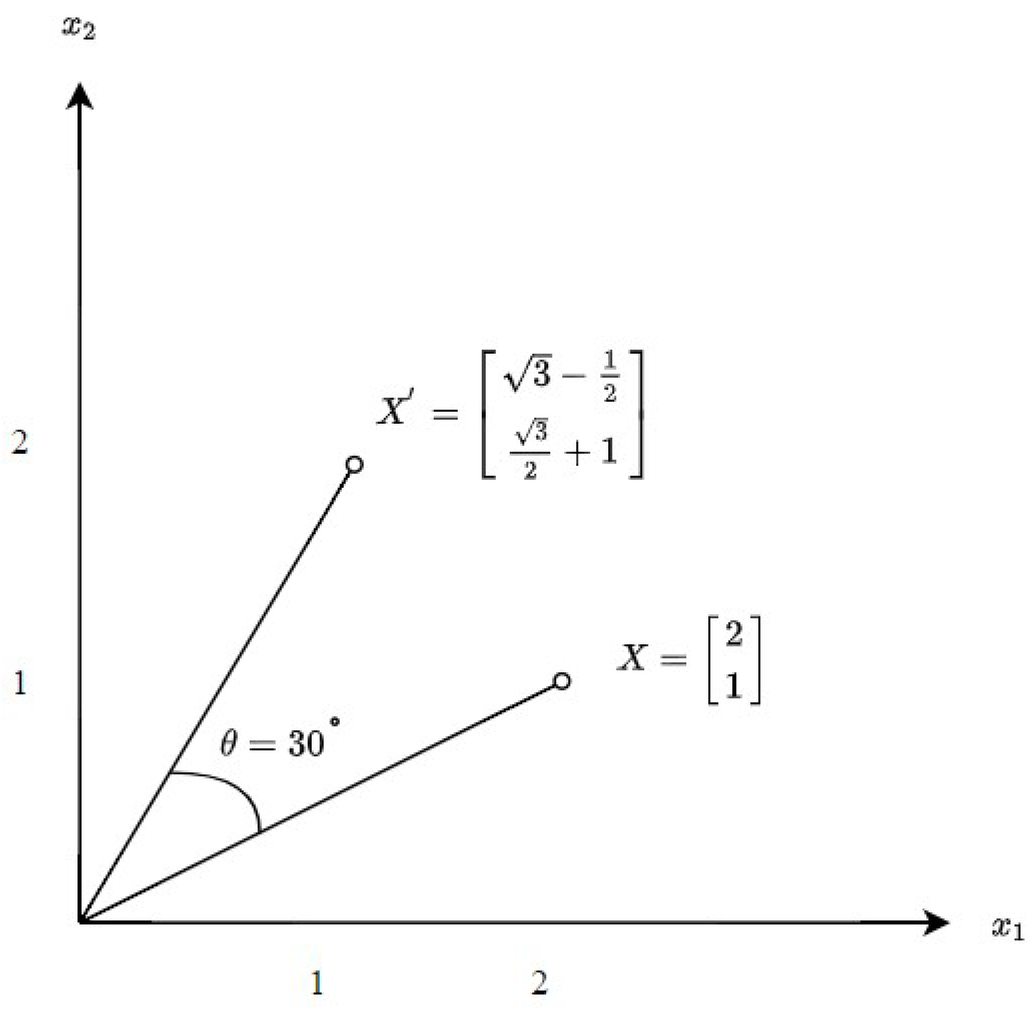
به طوری که $\mathbf{a}^\top_i \in \mathbb{R}^n$ بردار سطری نشان دهنده سطر i ام ماتریس A است. سپس ضرب ماتریس-بردار Ax را به کمک ضرب نقطه ای $\mathbf{a}^\top_i \mathbf{x}$ به صورت زیر تعریف می کنیم:

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a}_1^\top \mathbf{x} \\ \mathbf{a}_2^\top \mathbf{x} \\ \vdots \\ \mathbf{a}_m^\top \mathbf{x} \end{bmatrix}.$$

برای شهود بیشتر می توانید به این ضرب به چشم یک تبدیل از فضای $\mathbf{R}^n$ به $\mathbf{R}^m$ نگاه کنید. مثلا اگر $m = n = 2$ باشد و قرار دهیم

$$A = \begin{bmatrix} cos(30) & -sin(30) \\ sin(30) & cos(30) \end{bmatrix},\ X = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

در این صورت بردار $X' = AX$ در صفحه مختصات دو بعدی 30 درجه دوران یافته بردار X در جهت پادساعتگرد است.



در ادامه انجام این ضرب در پایتون را میبینیم:

```
A.shape, x.shape, A@x
```

```
((2, 3), (3,), array([ 5., 14.], dtype=float32))
```

اکنون آماده ایم تا ضرب ماتریس-ماتریس را معرفی کنیم. ابتدا قرار دهیم:

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{bmatrix},\quad \mathbf{B}=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{km} \end{bmatrix}.$$

سپس با نماد گذاری مشابه قبل داریم:

$$\mathbf{A}=\begin{bmatrix} \mathbf{a}^\top_{1} \\ \mathbf{a}^\top_{2} \\ \vdots \\ \mathbf{a}^\top_n \end{bmatrix},\quad \mathbf{B}=\begin{bmatrix} \mathbf{b}_{1} & \mathbf{b}_{2} & \cdots & \mathbf{b}_{m} \end{bmatrix}.$$

سپس می توانیم ماتریس C که $n \times m$ است را به صورت زیر تشکیل دهیم:

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} \mathbf{a}^\top_{1} \\ \mathbf{a}^\top_{2} \\ \vdots \\ \mathbf{a}^\top_n \end{bmatrix} \begin{bmatrix} \mathbf{b}_{1} & \mathbf{b}_{2} & \cdots & \mathbf{b}_{m} \end{bmatrix} = \begin{bmatrix} \mathbf{a}^\top_{1} \mathbf{b}_1 & \mathbf{a}^\top_{1}\mathbf{b}_2& \cdots & \mathbf{a}^\top_{1} \mathbf{b}_m \\ \mathbf{a}^\top_{2}\mathbf{b}_1 & \mathbf{a}^\top_{2} \mathbf{b}_2 & \cdots & \mathbf{a}^\top_{2} \mathbf{b}_m \\ \vdots & \vdots & \ddots &\vdots\\ \mathbf{a}^\top_{n} \mathbf{b}_1 & \mathbf{a}^\top_{n}\mathbf{b}_2& \cdots& \mathbf{a}^\top_{n} \mathbf{b}_m \end{bmatrix}.$$

در ادامه انجام این ضرب در پایتون را میبینیم:

```
B = np.ones([3, 4])
A@B
```

```
array([[ 3.,  3.,  3.,  3.],
       [12., 12., 12., 12.]])
```

5.1.نرم‌های برداری

فضای متریک یا متری S به مجموعه ای گفته می شود که مفهومی از فاصله با تابع $d : S \times S \to \mathbf{R}$ با شرایط زیر بین اعضای آن تعریف شده باشد:

$(x, y, z \in S)$

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \Leftrightarrow x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$

هر تابع به صورت $|| \cdot || : \mathbf{R}^n \to \mathbf{R}^+$ که در خواص سه گانه

1. $\|X\| = 0 \Leftrightarrow X = 0 \qquad \forall X \in \mathbf{R}^n$
2. $\|\alpha X\| = |\alpha| \|X\| \qquad \forall \alpha \in \mathbf{R}$
3. $\|X + Y\| \leq \|X\| + \|Y\| \qquad \forall X, Y \in \mathbf{R}^n$

صدق کند یک **نرم برداری** روی فضای برداری $\mathbf{R}^n$ نامیده می شود.

یکی از نرم های معروف نرم ℓ_2 است که به صورت زیر تعریف می شود:

$$\|\mathbf{X}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \langle \mathbf{X}, \mathbf{X} \rangle$$

یکی دیگر از نرم های معروف نرم ℓ_1 است که به صورت زیر تعریف می شود:

$$\|\mathbf{X}\|_1 = \sum_{i=1}^{n} |x_i| \, .$$

قضیه فیثاغورس) اگر X بر Y عمود باشد یا به طور معادل $\langle X, Y \rangle = 0$ آنگاه

$$\|\mathbf{X} + \mathbf{Y}\|_2^2 = \|\mathbf{X}\|_2^2 + \|\mathbf{Y}\|_2^2$$

نامساوی کشی-شوارتز)

$$|\langle X, Y \rangle| \leq \|\mathbf{X}\|_2 \|\mathbf{Y}\|_2$$

در کاربردهایی از یادگیری ماشین و یادگیری عمیق به دنبال بیشینه کردن یا کمینه کردن فاصله بازنمایی‌هایی از دادهایمان هستیم مثلا کمینه کردن فاصله بین بازنمایی تصاویر یک فرد به ما در تشخیص چهره وی کمک می کند که در این موارد نرم‌ها به عنوان معیاری برای توصیف فاصله بردارها به ما کمک می کنند.

تابع $|| . || : \mathbf{R}^{n \times n} \longrightarrow \mathbf{R}^+$ را یک نرم ماتریسی گوییم هرگاه در شرایط ۳ گانه زیر صدق کند.

1. $\|A\| \geq 0, \|A\| = 0 \iff A = 0 \;\; \forall A \in \mathbf{R}^{n \times n}$
2. $\|\alpha A\| = |\alpha| \|A\| A \;\; \forall \alpha \in \mathbf{R}$
3. $\|A + B\| \leq \|A\| + \|B\| \;\; \forall A, B \in \mathbf{R}^{n \times n}$

توجه کنید خاصیت سوم همان نامساوی مثلث است که البته گاهی اوقات به آن سازگاری جمعی هم گفته می‌شود.

هر تابع || . || که در خواص سه گانه بالا صدق کند یک نرم ماتریس خواهد بود. البته در کاربردها و آنالیز همگرایی روش‌ها اغلب لازم است یک نرم ماتریسی خاصیت اضافی دیگری به صورت $(4) \quad \|AB\| \leq \|A\| \|B\|$ داشته باشد که به آن خاصیت ضربی

گفته می‌شود.

بنابراین یک نرم ممکن است فقط در ۳ خاصیت گفته شده صدق کن و در خاصیت چهارم صدق نکند . اما همانطور که گفته شد اگر نرم ماتریسی که با آن کار می کنیم در خاصیت ضربی هم صدق کند بسیار مناسب و کاربردی تر خواهد بود.

اولین نرمی که ظاهراً تعمیم نرم ۲ برداری برای ماتریس‌هاست نرم فربنیوس است که به صورت $\|.\|_F$ نمایش داده می‌شود و برابر است با جذر مجموع مربعات درایه‌های ماتریس. به راحتی می‌توان نشان داد که هر ۴ شکلی که برای نرم فربنیوس در ادامه آورده شده‌اند باهم معادل می‌باشند.

1. فرض کنید A ماتریسی $n \times n$ باشد آنگاه:

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

2. فرض کنید $A = (a_{ij})$ آنگاه:

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

3. فرض کنید نمایش ستونی A به صورت $A = [a_1 \ \ a_2 \ \ \ldots \ \ a_n]$ باشد، آنگاه:

$$\|A\|_F = \sqrt{\sum_{i=1}^{n} \|a_i\|_2^2}$$

4. فرض کنید نمایش سطری A به صورت $A = \begin{bmatrix} a'_1 \\ \vdots \\ a'_n \end{bmatrix}$ باشد آنگاه:

$$\|A\|_F = \sqrt{\sum_{i=1}^{n} \|a'_i\|_2^2}$$

دسته‌ای دیگر از نرم‌های ماتریسی که بر اساس یک نرم برداری به صورت

$$\|A\| = \max_{X \neq 0} \frac{\|AX\|}{\|X\|}$$

ساخته می‌شوند را نرم القایی می‌نامیم. توجه کنید که دو نرم سمت راست رابطه‌ی فوق نرم برداری بوده و نرم سمت چپ رابطه بیانگر یک نرم ماتریسی است.

مثال) با استفاده از تعریف، مقدار $\|A\|_1$ را برای ماتریس $A = \begin{bmatrix} 0 & -1 \\ 2 & 0 \end{bmatrix}$ محاسبه کنید.

حل) فرض کنید $X \in \mathbf{R}^2$ طوری باشد که $\|X\|_1 = 1$ یعنی اگر $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ آنگاه $|x_1 \mid + \mid x_2 \mid = 1$ از طرفی داریم:

$$AX = \begin{bmatrix} 0 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ 2x_1 \end{bmatrix}$$

بنابراین داریم : $\|AX\|_1 = \left\| \begin{bmatrix} -x_2 \\ 2x_1 \end{bmatrix} \right\|_1 = \mid x_2 \mid +2 \mid x_1|$ لذا داریم :

$$\|A\|_1 = \max_{\|X\|_1 = 1} \|AX\|_1 = \max_{|x_1|+|x_2|=1} \{\mid x_2 \mid +2 \mid x_1 \mid\}$$

مساله اخیر یک مساله بهینه‌سازی (ماکزیمم) غیر خطی (به دلیل وجود قدر مطلق در تابع هدف $|x_2| + 2|x_1|$) مقید است. در واقع یک قید غیر خطی به صورت $|x_1| + |x_2| = 1$ مفروض است.

$$\max |x_2| + 2|x_1|$$
$$s.t$$
$$|x_1| + |x_2| = 1$$

که حل آن را به خواننده واگذار می کنیم.

قضیه) هرگاه A یک ماتریس حقیقی $n \times n$ باشد، آنگاه برای نرم‌های ماتریس القایی داریم:

$$(1) \quad ||A||_1 = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$$

$$(2) \quad ||A||_\infty = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|$$

6.1.مباحث تکمیلی

مجموعه بردارهای $\{v_1, v_2, \ldots, v_n\}$ در $\mathbb{R}^n$ متعامد است اگر :

$$v_i^T v_j = 0 \qquad , i \neq j$$

و بعلاوه می‌گوییم متعامد یکّه‌اند اگر :

$$v_i^T v_i = 1 \qquad , i = 1, 2, 3, \ldots, m$$

فرض کنید مجموعه $\{v_1, v_2, \ldots, v_m\}$ یک پایه برای $\mathbb{R}^n$ باشد. چنانچه مجموعه این بردارها متعامد یکّه باشد گوییم این پایه یک پایه یکّه متعامد برای $\mathbb{R}^n$ است.

مثال) فرض کنید مجموعه $\{v_1, v_2, v_3\}$ یک مجموعه یکّه متعامد برای $\mathbb{R}^3$ باشد. آنگاه بنا به تعریف پایه بودن هر عضو دلخواه $\mathbb{R}^3$ مثل u را می‌توان برحسب ترکیب خطی این بردارها نوشت. یعنی

$$u = c_1 v_1 + c_2 v_2 + c_3 v_3$$

ضرایب c_iها را محاسبه کنید.

حل) با ضرب طرفین رابطه‌ی فوق از سمت چپ در v_1^T داریم

$$v_1^T u = c_1 v_1^T v_1 + c_2 v_1^T v_2 + c_3 v_1^T v_3$$

چون مجموعه‌ی داده شده متعامد یکّه است پس داریم

$$v_1^T v_1 = 1 \ , \ v_1^T v_2 = 0 \ , \ v_1^T v_3 = 0$$

لذا با جایگذاری مقادیر فوق داریم

$$v_1^T u = c_1 \times 1 + c_2 \times 0 + c_3 \times 0 = c_1$$

پس ضریب c_1 به صورت $c_1 = v_1^T u$ حاصل می‌شود.

برای بدست آوردن v_2 کافی‌ست طرفین را از چپ در v_2^T ضرب کنیم، بنابراین داریم

$$v_2^T u = c_1 \underbrace{v_2^T v_1}_{0} + c_2 \underbrace{v_2^T v_2}_{1} + c_3 \underbrace{v_2^T v_3}_{0}$$

لذا $c_2 = v_2^T u$ کامل می‌شود، به طور مشابه $c_3 = v_3^T u$ کامل می‌شود.

فرض کنید بردارهای متعامد و یکّه v_1 و v_2 به صورت زیر داده شده‌اند.

$$v_1 = \frac{1}{7}\begin{bmatrix} 3 \\ -6 \\ 2 \end{bmatrix} \ , \ v_2 = \frac{2}{7}\begin{bmatrix} 1 \\ \frac{3}{2} \\ 3 \end{bmatrix}$$

آن‌ها را به عنوان ستون یک ماتریس مثل Q در نظر می‌گیریم.

$$Q = [\, v_1, \quad v_2 \,] = \begin{bmatrix} \frac{3}{7} & \frac{2}{7} \\ -\frac{6}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{6}{7} \end{bmatrix}_{3\times 2}$$

آنگاه همواره چنین ماتریسی خاصیت زیر را دارد:

$$Q^T Q = I$$

زیرا

$$Q^T Q = \begin{bmatrix} \frac{3}{7} & -\frac{6}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{3}{7} & \frac{6}{7} \end{bmatrix} \begin{bmatrix} \frac{3}{7} & \frac{2}{7} \\ -\frac{6}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{6}{7} \end{bmatrix} = \begin{bmatrix} \frac{9}{49} + \frac{36}{49} + \frac{4}{49} & \frac{6}{49} - \frac{18}{49} + \frac{12}{49} \\ \frac{6}{49} - \frac{18}{49} + \frac{12}{49} & \frac{4}{49} + \frac{9}{49} + \frac{36}{49} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

$\Leftarrow$ درواقع به چنین ماتریس‌هایی که ستون‌هایش متعامدند، **ایزومتری** گوییم.

توجه کنید که اگر ماتریس Q که توسط بردارهای متعامد یکّه v_i ساخته می‌شود، مربعی باشد علاوه بر خاصیت $Q^T Q = I$ دارای خاصیت $QQ^T = I$ نیز هست که در این حالت به آن یک ماتریس **متعامد** می‌گوییم.

برای مثال اگر بردارهای متعامد یکّه زیر را داشته باشیم

$$v_1 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \ , \ v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \ , \ v_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

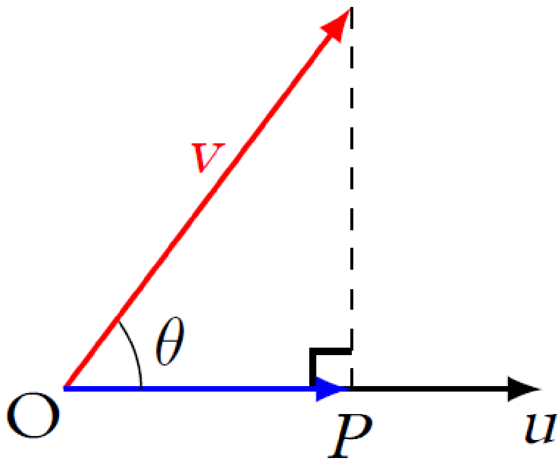
آنگاه ماتریس ساخته شده با بردارهای v_1 ، v_2 و v_3 مربعی و به صورت زیر است:

$$Q = [v_1, v_2, v_3] = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -1 & 0 & 0 \end{bmatrix}_{3\times 3}$$

و با محاسباتی ساده می‌توان دید که Q در $Q^T Q = I$ و $QQ^T = I$ صدق می‌کند، لذا ماتریس Q یک ماتریس متعامد است.

لازم است که بیان کنیم در یک ماتریس ایزومتری تنها ستون‌ها بر هم متعامد یکّه‌اند امّا در یک ماتریس متعامد هم ستون‌ها و هم سطرها بر هم عمودند.

فرض کنید دو بردار u و v داده شده‌اند و بردار v با بردار u زاویه θ بسازد.(شکل زیر را ببینید)



$$\frac{\pi}{۲} > \theta > ۰ \,,$$

از بردار v عمودی بر بردار u رسم می‌کنیم و محل تقاطع را P نام‌گذاری می کنیم. آنگاه بردار OP را **تصویر** v بر u می‌نامیم.

برای بدست آوردن بردار OP برحسب u و v به صورت زیر عمل می‌کنیم:

$$\triangle OAP : \cos(\theta) = \frac{\text{طول ضلع مجاور}}{\text{طول وتر}} = \frac{\|OP\|_2}{\|v\|_2}$$
$$\rightarrow \|OP\|_2 = \|v\|_2 \cos(\theta) \qquad (1)$$

از طرفی برای ضرب داخلی u و v داریم:

$$\langle u, v\rangle = \langle v, u\rangle = \|v\|_2 \|u\|_2 \cos(\theta) \qquad (2)$$

از (1) و (2) داریم

$$\|OP\|_2 = \frac{\langle v, u\rangle}{\|u\|_2}$$

اکنون برای بدست آوردن بردار تصویر OP کافی است بردار یکّه و همجهت $\frac{u}{\|u\|_2}$ را در $\|OP\|_2$ ضرب کنیم:

$$OP = \frac{\langle v, u\rangle}{\|u\|_2} . \frac{u}{\|u\|_2} = \frac{\langle v, u\rangle}{\|u\|_2^2} . u$$

معمولا بردار فوق را با نماد $proj_u(v)$ نمایش می‌دهند.

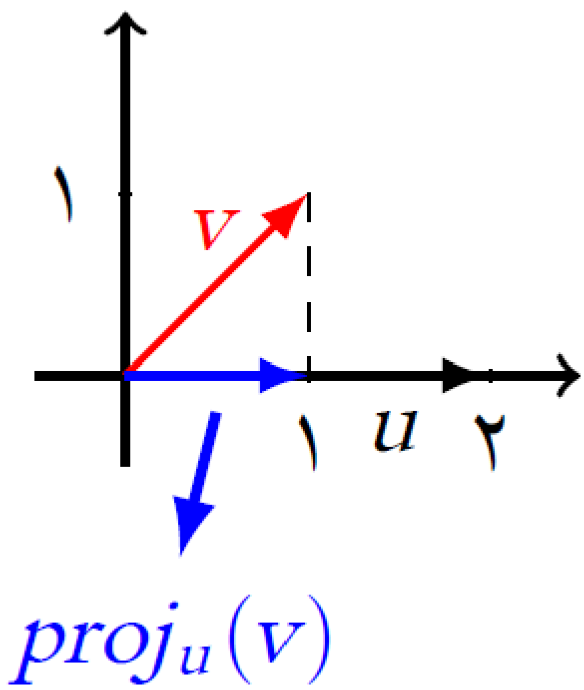
$$proj_u(v) = \frac{\langle v, u\rangle}{\|u\|_2^2} . u$$

مثال) تصویر بردار $v = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ را بر بردار $u = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ بدست آورید.

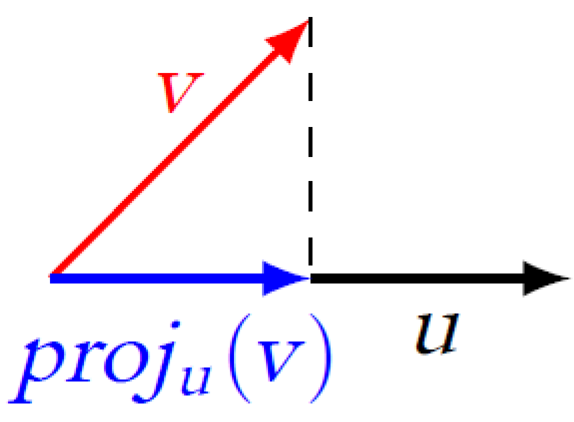
حل) داریم

$$\langle v, u\rangle = v^T u = [\,1 \quad 1\,] \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2$$
$$\|u\|_2^2 = 0^2 + 2^2 = 4$$
$$proj_u(v) = \frac{\langle v, u\rangle}{\|u\|_2^2} . u = \frac{2}{4} u = \frac{1}{2} u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

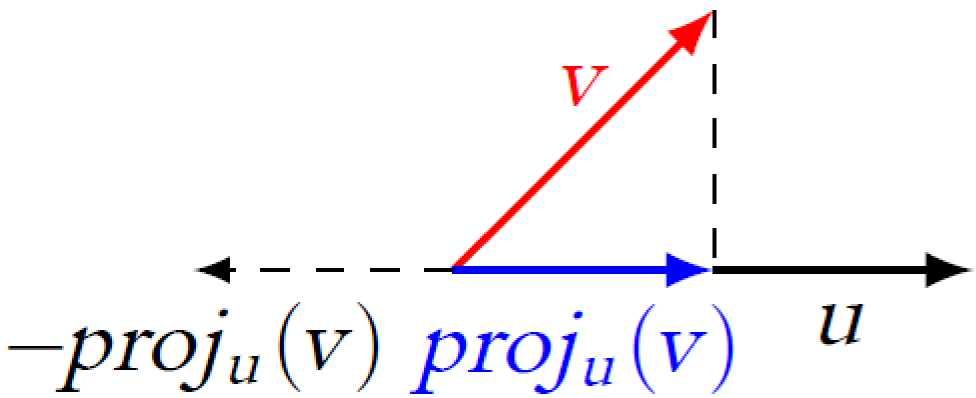
نتیجه کاملا با شکل زیر مطابقت دارد.



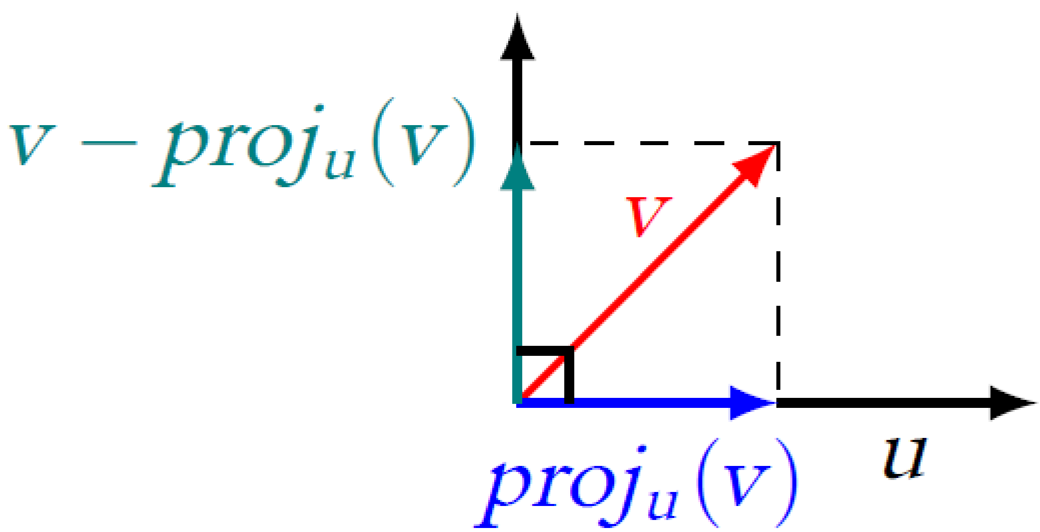
فرض کنید دو بردار u و v داده شده‌اند. به تصویر بردار v بر u دقت کنید.



بردار $-proj_u(v)$ به صورت زیر است:



لذا جمع دو بردار v و $-proj_u(v)$ به صورت زیر است:



این از لحاظ هندسی نشان می‌دهد که بردار $v - proj_u(v)$ بر بردار u عمود است. البته این نتیجه را می‌توان به صورت جبری نیز اثبات کرد؛ زیرا

$$\begin{aligned}\langle v - proj_u(v), u\rangle &= \langle u, v - proj_u(v)\rangle = u^T(v - proj_u(v)) \\ &= u^Tv - u^Tproj_u(v) = u^Tv - u^T(\frac{\langle v, u\rangle}{\|u\|_2^2}).\,u \\ &= u^Tv - u^T(\frac{\langle u, v\rangle}{\|u\|_2^2}).\,u = u^Tv - u^T(\frac{u^Tv}{\|u\|_2^2}).\,u \\ &= u^Tv - (\frac{u^Tv}{\|u\|_2^2})u^Tu\ ;\ u^Tu = \|u\|_2^2 \\ &= u^Tv - (\frac{u^Tv}{\|u\|_2^2}).\,\|u\|_2^2 = u^Tv - u^Tv = 0\end{aligned}$$

بنابر این ثابت کردیم که

$$v - proj_u(v) \perp u$$

فرض کنید A ماتریسی $n \times n$ باشد. آنگاه (λ, x) یک **جفت ویژه** ماتریس A نامیده می‌شوند اگر

$$AX = \lambda X \qquad (1)$$

توجه کنید X برداری غیرصفر است و λ یک اسکالر است. از تعریف (1) داریم

$$AX - \lambda X = 0 \ \Rightarrow\ AX - \lambda I_n X = 0 \ \Rightarrow\ (A - \lambda I_n)X = 0$$

چون فرض کرده ایم $X \neq 0$ پس باید ماتریس $A - \lambda I_n$ منفرد باشد یعنی دترمینان آن مساوی صفر باشد زیرا اگر ناصفر باشد آنگاه وارون پذیر بوده و

$$(A - \lambda I_n)X = 0 \ \Rightarrow\ (A - \lambda I_n)^{-1} \times (A - \lambda I_n)X = (A - \lambda I_n)^{-1} \times 0 = 0 \Rightarrow IX = 0 \Rightarrow X = 0$$

که تناقض است پس ماتریس $(A - \lambda I_n)$ باید منفرد باشد یعنی $det(A - \lambda I_n) = 0$ می توان دید که $P_A(\lambda) = det(A - \lambda I_n)$ یک چندجمله ای برحسب λ و دقیقا درجه n است و آن را **چندجمله ای مشخصه** ماتریس A می نامند. بنابر (1) ریشه‌های $P_A(\lambda)$ همان **مقادیر ویژه** A هستند لذا بنابر قضیه اساسی جبر ماتریس A دارای n مقدار ویژه است.

قضیه اساسی جبر) هر چندجمله ای درجه n در اعداد مختلط دارای n ریشه با احتساب تکرر است.

$$p(x) = a(x - \lambda_1)^{t_1}(x - \lambda_2)^{t_2} \cdots (x - \lambda_k)^{t_k}$$

مثال) جفت ویژه ماتریس زیر را محاسبه کنید.

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

حل) ابتدا چندجمله ای مشخصه را محاسبه می کنیم

$$P_A(\lambda) = det(A - \lambda I_2)$$

$$A - \lambda I_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}$$

لذا

$$P_A(\lambda) = det(A - \lambda I_2) = det\left(\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}\right)$$
$$= (2-\lambda)(2-\lambda) - (1)(1) = \lambda^2 - 4\lambda + 4 - 1$$
$$= \lambda^2 - 4\lambda + 3$$
$$P_A(\lambda) = 0 \Rightarrow \lambda^2 - 4\lambda + 3 = 0 \Rightarrow (\lambda - 1)(\lambda - 3) = 0 \Rightarrow \lambda_1 = 1, \lambda_2 = 3$$

برای بدست آوردن بردار ویژه متناظر با $\lambda_1 = 1$ داریم

$$(A - \lambda_1 I_2)X_1 = 0 \Rightarrow \begin{bmatrix} 2-\lambda_1 & 1 \\ 1 & 2-\lambda_1 \end{bmatrix} X_1 = 0 \Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} X_1 = 0 \Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\Rightarrow \begin{cases} x_1 + x_2 = 0 \\ x_1 + x_2 = 0 \end{cases} \Rightarrow \; x_1 = -x_2$$

پس داریم

$$X_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

چون باید $X_1 \neq 0$ کافی است انتخاب کنیم به طور دلخواه $x_2 = 1$ پس بردار ویژه X_1 به صورت زیر حاصل می شود:

$$X = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

برای بردار ویژه متناظر با $\lambda_2 = 3$ داریم

$$(A - \lambda_2 I_2)X_2 = 0 \Rightarrow \begin{bmatrix} 2-\lambda_2 & 1 \\ 1 & 2-\lambda_2 \end{bmatrix} X_2 = 0 \Rightarrow \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} X_2 = 0 \Rightarrow \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\Rightarrow \begin{cases} -x_1 + x_2 = 0 \\ x_1 - x_2 = 0 \end{cases} \Rightarrow \; x_1 = x_2$$

پس داریم

$$X_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

چون باید $X_2 \neq 0$ کافی است انتخاب کنیم به طور دلخواه $x_2 = 1$ پس بردار ویژه X_2 به صورت زیر حاصل می شود:

$$X_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

نکته)

$$trace(A) = \sum_i \lambda_i, \qquad det(A) = \prod_i \lambda_i, \qquad trace(A^k) = \sum_i \lambda_i^k$$

ماتریس A را **قطری شدنی** گویند هرگاه ماتریس نامنفرد P موجود باشد که

$$A = PDP^{-1}$$

که در این صورت گوییم A با یک ماتریس قطری D متشابه است. وقتی ماتریس A را بتوان به صورت فوق تجزیه نمود آنگاه می‌توان از فوایدی بهره برد. برای مثال عناصر قطری D همان مقادیر ویژه‌ی A خواهند بود.

به علاوه توان k- ام ماتریس A را به راحتی از رابطه‌ي

$$A^k = PDP^{-1}$$

به دست می‌آید. توجه کنید که توان k- ام ماتریس قطری

$$D^k = \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & d_n \end{bmatrix}$$

به راحتی از

$$D^k = \begin{bmatrix} d_1^k & 0 & 0 & 0 \\ 0 & d_2^k & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & d_n^k \end{bmatrix}$$

محاسبه می‌شود. اکنون ماتریس زیر را در نظر بگیرید

$$A = \begin{pmatrix} 3 & 2 & 2 \\ 1 & 2 & 2 \\ -1 & -1 & 0 \end{pmatrix}$$

با محاسبه‌ی جفت ویژه‌های A داریم.

$$\lambda_1 = 1, \;\; \lambda_2 = 2, \;\; \lambda_3 = 2, \;\; X_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \;\; X_2 = X_3 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

چون A دارای سه بردار ویژه‌ی مستقل خطی نیست پس قطری شدنی نیست. بنابراین وجود تجزیه‌ای به صورت $A = PDP^{-1}$ ممکن نیست و این محدودیت‌هایی در کار کردن با ماتریس A (در این مثال) می‌تواند به وجود آورد. با توجه به اینکه در عمل ممکن است با ماتریس‌هایی سروکار داشته باشیم که قطری شدنی نباشند یعنی دارای تجزیه‌ای به شکل $A = PDP^{-1}$ نباشند پس معقول است که به دنبال یک تجزیه‌ی جایگزین برای A باشیم که احتمالا از خواص خوبی مانند آنچه در تجزیه‌ی قطری شدنی دارد، دارا باشد. به علاوه چنین تجزیه‌ی جایگزینی می‌بایست برای هر ماتریس دلخواه موجود باشد و از محدودیت‌های تجزیه قطری شدنی برخوردار نباشد.

عموما چنین تجزیه‌ای جایگزینی برای A وجود دارد که به صورت

$$A = U\Sigma V^T \tag{1}$$

بیان می‌شود که در آن U, V ماتریس‌هایی متعامدند یعنی

$$UU^T = U^TU = I, \qquad VV^T = V^TV = I$$

و Σ ماتریسی قطری است. مثلا برای ماتریس همین مثال داریم

$$\begin{bmatrix} 3 & 2 & 2 \\ 1 & 2 & 2 \\ -1 & -1 & 0 \end{bmatrix} = \begin{bmatrix} -0.7984 & 0.4400 & 0.4110 \\ -0.5545 & -0.8033 & -0.2172 \\ 0.2346 & -0.4013 & 0.8854 \end{bmatrix} \times \begin{bmatrix} 5.1096 & 0 & 0 \\ 0 & 1.2152 & 0 \\ 0 & 0 & 0.6442 \end{bmatrix} \times \begin{bmatrix} -0.6232 & 0.7554 & 0.2023 \\ -0.5755 & -0.2678 & -0.7727 \\ -0.5296 & -0.5980 & 0.6016 \end{bmatrix}^T$$

به علاوه چون V متعامد است پس $V^T = V^{-1}$ و لذا (1) را می‌توان نوشت

$$A = U\Sigma V^{-1}, \text{(ماتریس قطری است}\Sigma)$$

که شباهت فراونی با رابطه‌ی قطری شدنی $A = PDP^{-1}$ دارد. به تجزیه‌ی (1) **تجزیه مقادیر تکین (Singular Value Decomposition)** یا به اختصار تجزیه SVD ماتریس A می‌گوییم که صورت رسمی در ادامه تعریف می‌شود.

قضیه) فرض کنید A یک ماتریس حقیقی $m \times n$ باشد، آنگاه ماتریس‌های متعامد U و V وجود دارند به قسمی که

$$U^T AV = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} = \Sigma$$

که در آن Σ_1 یک ماتریس قطری نامنفرد است. عناصر قطری Σ همگی نامنفی هستند و می‌توانند به ترتیب ناصعودی مرتب شوند. تعداد عناصر قطری مخالف صفر Σ برابر رتبه ماتریس A است.

تعریف) عناصر قطری ماتریس Σ مقادیر تکین ماتریس A نامیده می‌شوند. اعداد $\sigma_1, \sigma_2, \ldots, \sigma_r$ مقادیر تکین مثبت A هستند.

تعریف) ستون‌های U بردارهای تکین چپ و ستون‌های V بردارهای تکین راست نامیده می‌شوند.

توجه) به تعداد $k = \min(m, n)$ مقدار تکین برای A وجود دارد. فرض کنید رتبه A برابر r باشد. پس r مقدار تکین مثبت وجود دارد. این‌ها ریشه‌های دوم مثبت مقادیر ویژه مخالف صفر $A^T A$ یا AA^T هستند. بقیه $(k - r)$ مقدار تکین، اگر $r < k$، صفر هستند. بنابراین مقادیر تکین منحصر به فرد هستند ولیکن بردارهای تکین منحصر به فرد نیستند.

مثال) مقادیر تکین ماتریس داده شده را حساب کنید.

$$A = \begin{bmatrix} 2 & -2 & 1 \\ 2 & 1 & 3 \end{bmatrix}, \qquad m = 2, n = 3$$

حل) چون $A^T A$ ماتریس 3×3 و AA^T ماتریس 2×2 است. پس منطقی است که جذر مقادیر ویژه‌ی ماتریس AA^T را بیابیم.

$$AA^\top = \begin{bmatrix} 9 & 5 \\ 5 & 14 \end{bmatrix} \rightarrow P(\lambda) = \lambda^2 - 23\lambda + 101$$

ریشه‌های چندجمله‌ای فوق به صورت

$$\lambda_1 = 17.0902, \quad \lambda_2 = 5.9098$$

هستند و از آنجا

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{17.0902} = 4.1340$$
$$\sigma_2 = \sqrt{\lambda_2} = \sqrt{5.9098} = 2.4310$$

مقادیر تکین A هستند. توجه کنید مطابق قضیه تعداد مقادیر تکین مثبت برابر رتبه A است پس در اینجا رتبه A برابر 2 خواهد بود.

قضیه) فرض کنید تجزیه SVD ماتریس $A \in \mathbb{R}^{m\times n}$ به صورت $A = U\Sigma V^T$ باشد آنگاه

$$AA^T U = U\tilde{\Sigma}, \qquad A^T AV = V\hat{\Sigma}$$

که در آن $\tilde{\Sigma}$ ماتریس $m \times m$ است به طوری که

$$\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & 0 \end{bmatrix}$$

و $\hat{\Sigma}$ ماتریسی $n \times n$ است به قسمی که

$$\hat{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & 0 \end{bmatrix}$$

و $\tilde{\Sigma}_1$ ماتریسی قطری $\tilde{\Sigma}_1 = diag(\sigma_1^2 \sigma_2^2, \ldots, \sigma_s^2)$ با $s = min\{m, n\}$ است.

نحوه‌ی محاسبه‌ی تجزیه SVD به صورت زیر می‌باشد :

1. ماتریس‌های $A^T A$ و AA^T را بسازید.
2. جفت ویژه‌ی این دو ماتریس را محاسبه کنید و ماتریس Σ را تشکیل دهید.

3. بردارهای ویژه AA^T را به عنوان ستون‌های U و بردارهای ویژه A^TA را به عنوان ستون‌های V در نظر بگیرید.
اکنون تجزیه SVD ماتریس A کامل شده است.

مثال) تجزیه SVD ماتریس داده شده را به دست آورید

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 0 & 3 \end{bmatrix}, \quad m = 3, n = 2$$

حل) گام1: ابتدا ماتریس‌های AA^T و A^TA را تشکیل می‌دهیم

$$AA^T = \begin{bmatrix} 2 & 1 & -3 \\ 1 & 5 & 3 \\ -3 & 3 & 9 \end{bmatrix}, \quad A^TA = \begin{bmatrix} 5 & 1 \\ 1 & 11 \end{bmatrix}$$

گام2: جفت ویژه‌های A^TA چنین اند:

$$\lambda_1 = 11/1623 \to v_1 = \begin{bmatrix} 0/1602 \\ 0/9871 \end{bmatrix}, \quad \lambda_2 = 4/8377 \to v_2 = \begin{bmatrix} -0/9871 \\ 0/1602 \end{bmatrix}$$

جفت ویژه‌های AA^T چنین اند:

$$\lambda_1 = 11/1623 \to u_1 = \begin{bmatrix} -0/2475 \\ 0/3913 \\ 0/8863 \end{bmatrix}, \quad \lambda_2 = 4/8377 \to u_2 = \begin{bmatrix} -0/2475 \\ 0/3913 \\ 0/8863 \end{bmatrix}$$

$$\lambda_3 = 0 \to u_3 = \begin{bmatrix} -0/8165 \\ 0/4082 \\ -0/4082 \end{bmatrix}$$

بعلاوه 2 مقدار تکین به صورت زیر موجود است:

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{11/1623} = 3/3410, \quad \sigma_2 = \sqrt{\lambda_2} = \sqrt{4/8377} = 2/1995$$

پس $r = 2$ و لذا Σ_1 ماتریسی $r \times r = 2 \times 2$ است

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} 3/3410 & 0 \\ 0 & 2/1995 \end{bmatrix}$$

در نتیجه ماتریس Σ ، 3×2 به صورت زیر است

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3/3410 & 0 \\ 0 & 2/1995 \\ 0 & 0 \end{bmatrix}$$

گام3: با توجه به بردارهای ویژه‌ی ماتریس‌های AA^T و A^TA ماتریس‌های U و V را به صورت زیر تشکیل می‌دهیم

$$U = [u_1, u_2, u_3] = \begin{bmatrix} -0/2475 & 0/5216 & -0/8165 \\ 0/3913 & 0/8247 & 0/4082 \\ 0/8863 & -0/2185 & -0/4082 \end{bmatrix}$$

$$V = [v_1, v_2] = \begin{bmatrix} 0/1602 & -0/9871 \\ 0/9871 & 0/1602 \end{bmatrix}$$

در نتیجه

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 0 & 3 \end{bmatrix} = U\Sigma V^T$$

$$= \begin{bmatrix} -0/2475 & 0/5216 & -0/8165 \\ 0/3913 & 0/8247 & 0/4082 \\ 0/8863 & -0/2185 & -0/4082 \end{bmatrix} \begin{bmatrix} 3/3410 & 0 \\ 0 & 2/1995 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0/1602 & -0/9871 \\ 0/9871 & 0/1602 \end{bmatrix}$$

منابع

1. جزوه جبرخطی عددی دکتر دهقان، دانشکده ریاضی و علوم کامپیوتر، دانشگاه صنعتی امیرکبیر
2. فصل دوم کتاب Dive Into Deep Learning
3. جزوه ریاضیات برای یادگیری ماشین دانشگاه Berkely