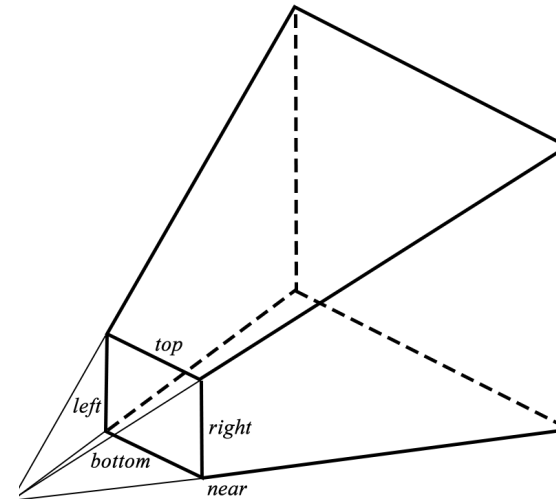


# مبانی جبر خطی

هادی عاشری



## مقدمه

◀ جبر خطی یکی از پایه‌های اساسی در یادگیری ماشین و یادگیری عمیق است و کاربردهای گسترده‌ای دارد:

◀ نمایش داده‌ها به صورت بردار، ماتریس، تنسور

◀ عملیات شبکه‌های عصبی با استفاده از تبدیل‌های خطی

◀ روش‌هایی مانند تجزیه مقادیر منفرد و تجزیه مقادیر ویژه برای کاهش ابعاد داده‌ها و انتخاب ویژگی‌ها بر پایه جبر خطی کار می‌کنند.

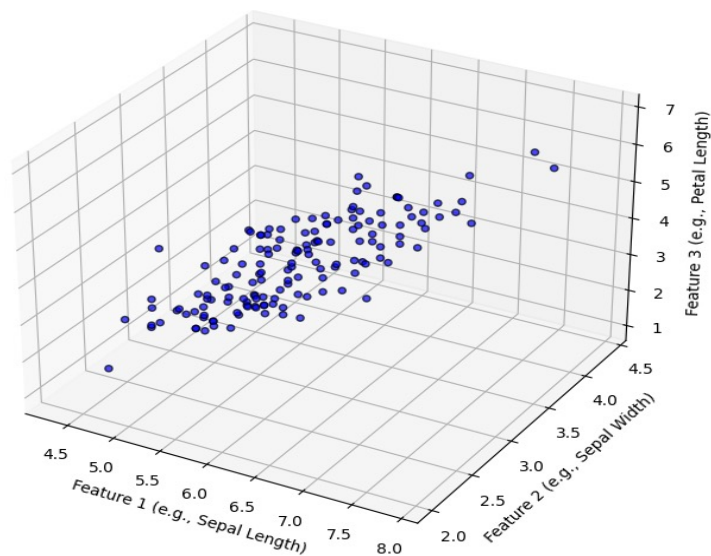
◀ الگوریتم‌های بهینه‌سازی مانند روش گرادیان نزولی برای تنظیم وزن‌ها از جبر خطی بهره می‌برند.

◀ تبدیل‌های ماتریسی مانند تبدیل فوریه برای فشرده‌سازی و تحلیل داده‌های تصویری استفاده می‌شوند.

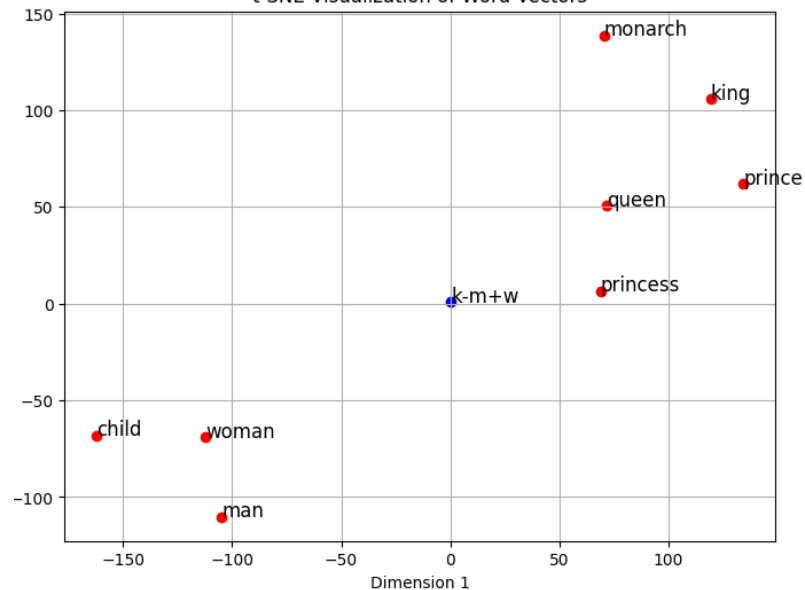


# جبر خطی: نمایش داده‌ها به صورت بردار

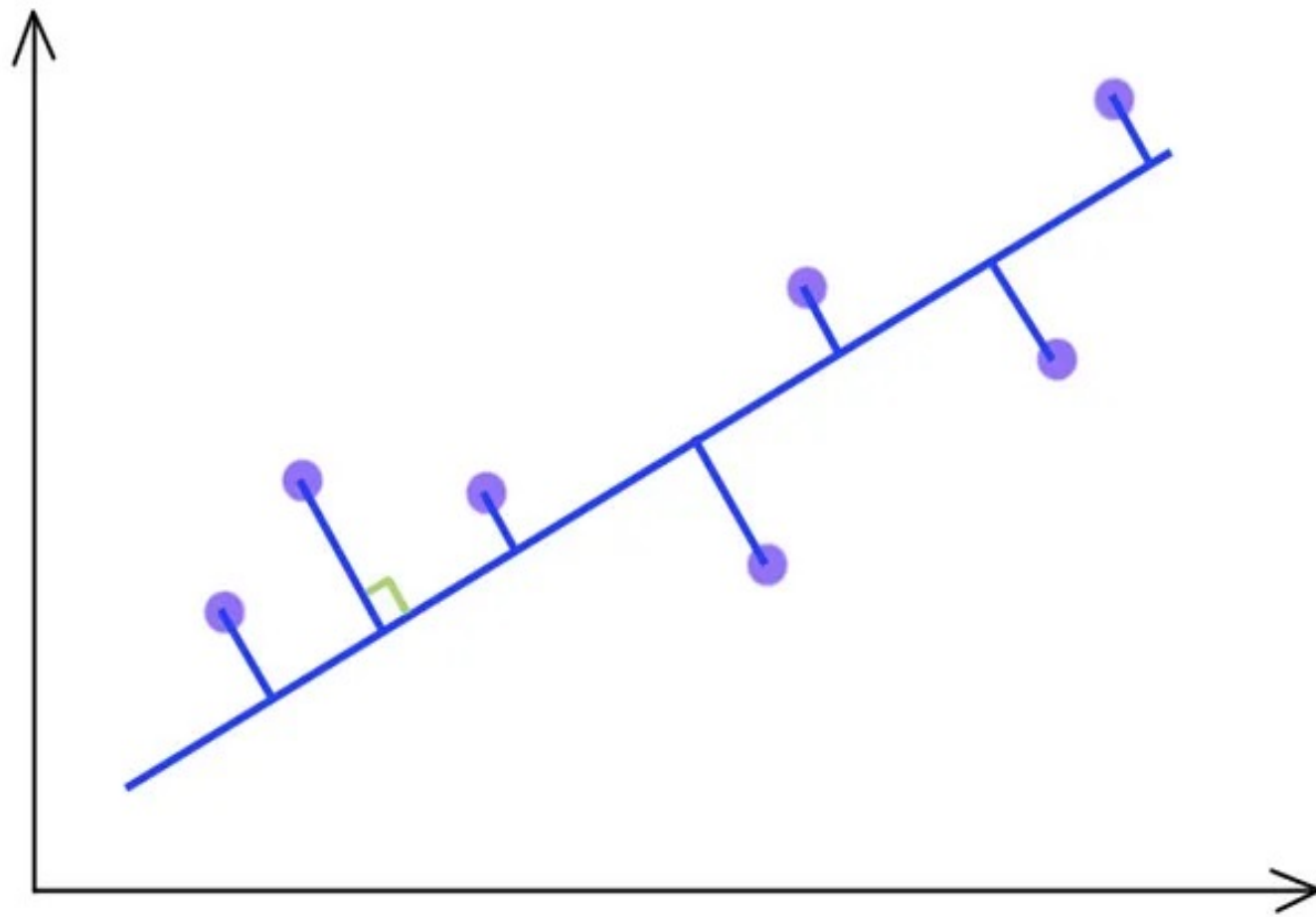
Visualization of Vectors as 3D Dots (First Three Features)



t-SNE Visualization of Word Vectors



## تجزیه مقادیر ویژه برای کاهش بعد



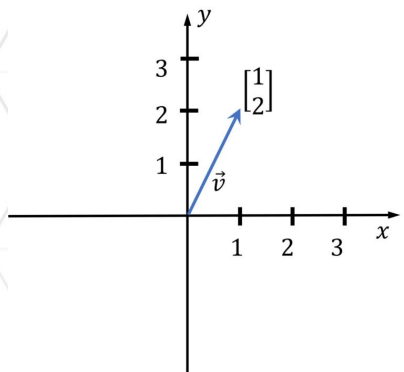
# بردارها

تعریف:

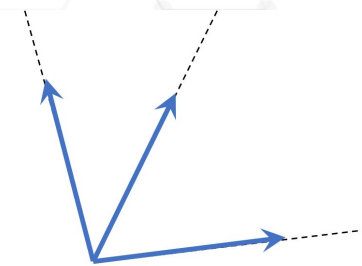
بردار یک موجودیت ریاضی است.

معمولاً به صورت آرایه‌ای از اعداد به شکل زیر نمایش داده می‌شود:

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$



بردار دارای جهت و اندازه است.



Direction



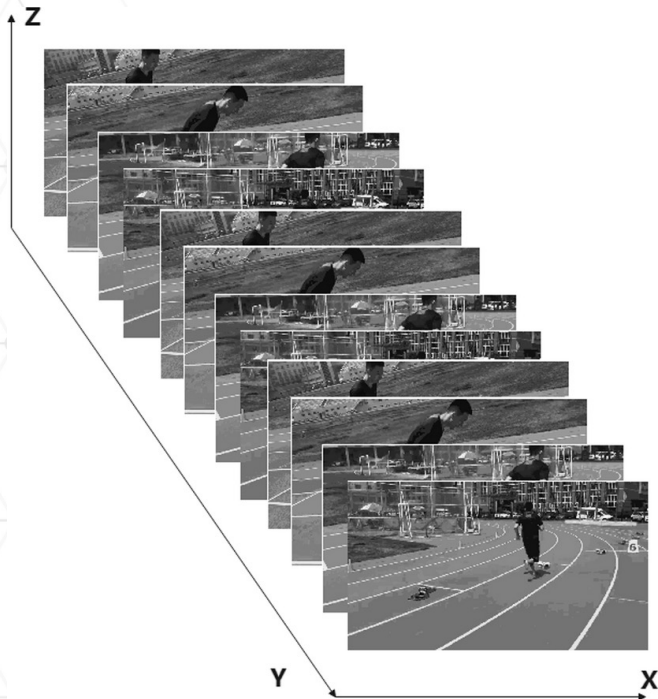
Length





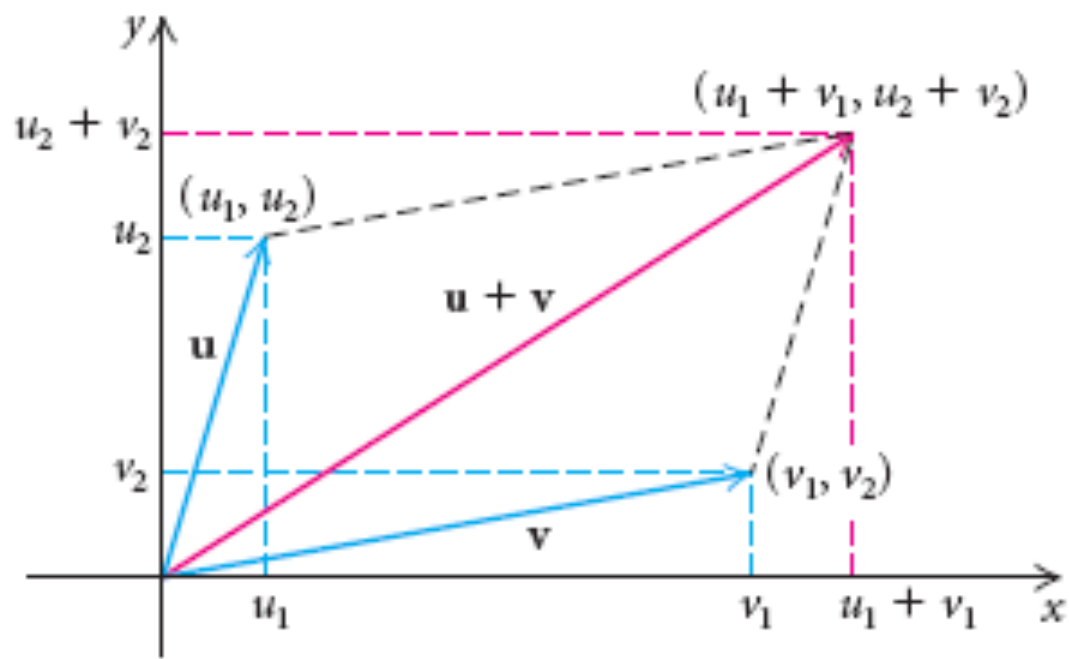
## بازنمایی ویدیو

◀ در بینایی ماشین، یک ویدیوی HD سیاه‌سفید یک ثانیه‌ای با نرخ فریم ۳۰ به‌عنوان دنباله‌ای ۳۰ تایی از بردارهای دوبعدی/ با ابعاد  $720 \times 1280$  (یک ماتریس سه بعدی) بازنمایی شود.



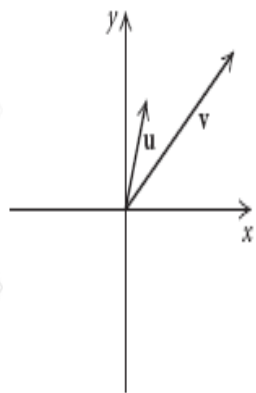
## عملیات جبری: جمع بردارها

$$\mathbf{u} + \mathbf{v} = [u_1 + v_1, u_2 + v_2, \dots, u_n + v_n]$$

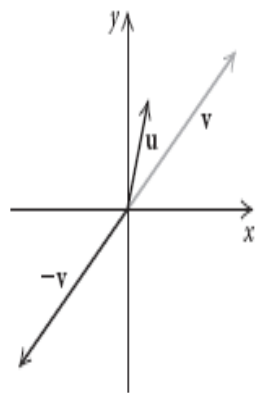


## عملیات جبری: تفریق بردارها

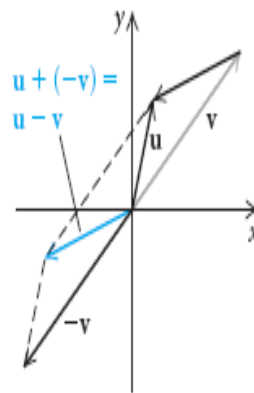
$$\mathbf{u} - \mathbf{v} = [u_1 - v_1, u_2 - v_2, \dots, u_n - v_n]$$



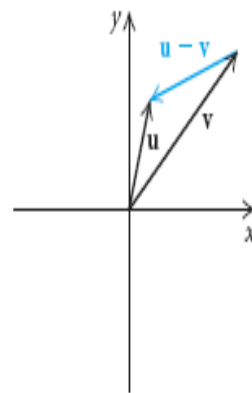
Sketch  $\mathbf{u}$  and  $\mathbf{v}$ .



Sketch  $-\mathbf{v}$ .



Sketch  $\mathbf{u} + (-\mathbf{v})$ , or  $\mathbf{u} - \mathbf{v}$ , using the parallelogram law.

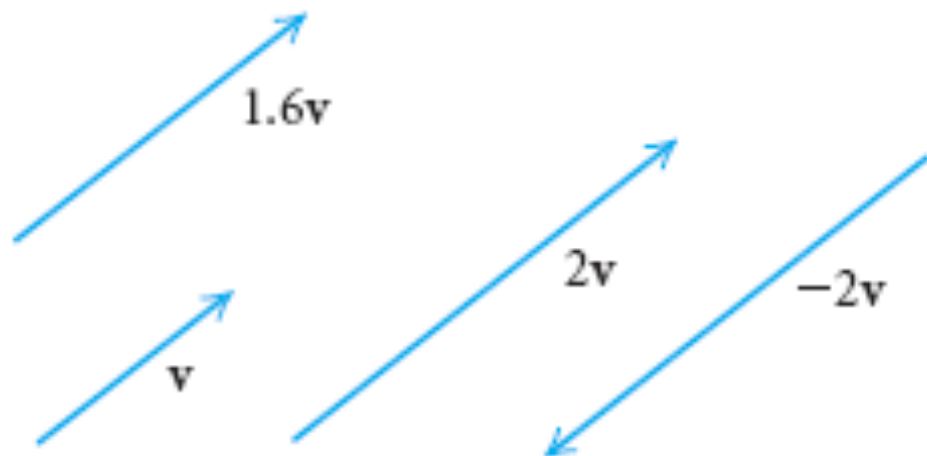


$\mathbf{u} - \mathbf{v}$  is the vector from the terminal point of  $\mathbf{v}$  to the terminal point of  $\mathbf{u}$ .



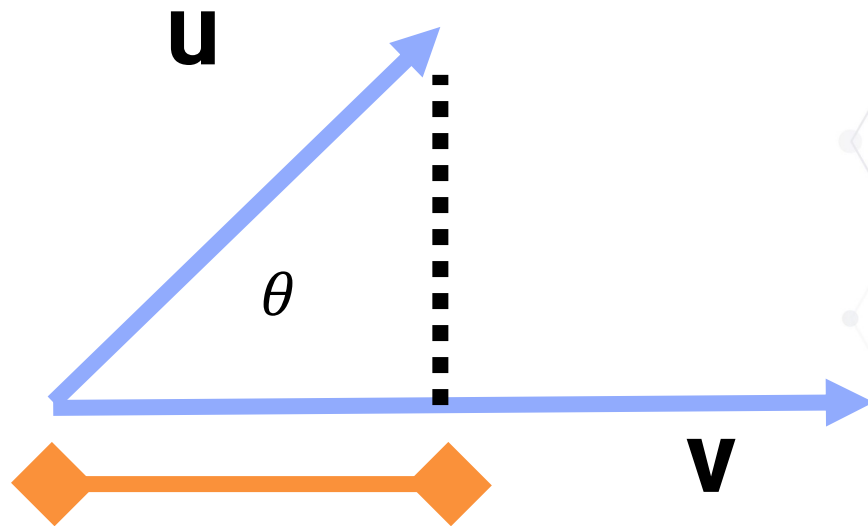
## عملیات جبری: ضرب اسکالر

$$cv = [cv_1, cv_2, \dots, cv_n]$$



## عملیات جبری: ضرب داخلی

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n = \|\mathbf{v}\| \|\mathbf{u}\| \cos \theta$$



## سنجش مشابهت بردارها

می‌توان از ضرب داخلی برای تعیین میزان مشابهت بین دو بردار ویژگی استفاده کرد.

$$u \cdot v = \|u\| \|v\| \cos \theta$$

تشابه کسینوسی:

$$\cos \theta = \frac{u \cdot v}{\|u\| \|v\|}$$



## شهود ضرب داخلی

ضرب داخلی یعنی اندازه‌ی هم‌جهتی دو بردار.

اگر بردارها کاملاً هم‌جهت باشند، ضرب داخلی‌شان بیشترین مقدار ممکن است.

اگر کاملاً عمود باشند، ضرب داخلی‌شان صفر است (چون هیچ هم‌جهتی ندارند).

اگر خلاف جهت باشند، ضرب داخلی‌شان منفی می‌شود (یعنی در جهت مخالف هم‌اند)





## ماتریس‌ها

ماتریس یک آرایه دوبعدی از اعداد است که در سطرها و ستون‌ها سازماندهی شده است.

معمولاً به صورت زیر نشان داده می‌شود:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = [a_{ij}]$$

در یادگیری ماشین می‌توان یک داده را با یک بردار ویژگی و مجموعه‌ای از داده‌ها را با یک ماتریس نیز نمایش داد.

تنسور تعمیم‌یافته‌ی ماتریس به بیش از دو بُعد یعنی یک آرایه چند بعدی است.



## فضای برداری

- ▶ یک فضای برداری (Vector Space) مجموعه‌ای از بردارها/اشیاء است که دو عمل جمع برداری و ضرب عددی (ضرب اسکالر) بر روی آن‌ها تعریف شده و مجموعه‌ای از اصول (آکسیوم‌ها) را برآورده می‌کند:
- ▶ بسته بودن نسبت به جمع: اگر  $u$  و  $v$  بردارهای فضای برداری باشند: آنگاه  $u + v$  نیز باید در فضای برداری وجود داشته باشد.
- ▶ بسته بودن نسبت به ضرب اسکالر: برای هر بردار  $u$  و هر عدد حقیقی  $c$ ، بردار  $cu$  نیز در فضای برداری وجود دارد.
- ▶ وجود بردار صفر: برداری که با هر برداری جمع شود همان بردار را به دست می‌دهد.
- ▶ وجود معکوس جمعی: برای هر بردار  $u$ ، قاعده  $u + (-u) = 0$  برقرار است.



## فضای برداری: مثال‌ها

◀ فضای  $\mathbb{R}^n$  مجموعه تمام بردارهای با  $n$  مولفه حقیقی

◀ مجموعه تمام چندجمله‌ای‌ها با درجه کمتر یا مساوی  $n$

◀ فضای ویژگی در یادگیری ماشین

◀ تصاویر دیجیتال که به صورت ماتریس‌های پیکسلی نمایش داده می‌شوند می‌توانند به تانسور تبدیل شوند.

◀ تبدیل کلمات به بردار و ایجاد فضای برداری کلمات با استفاده از الگوریتم‌هایی مانند Word2Vec، GloVe یا FastText



## عملیات روی ماتریس‌ها

◀ جمع ماتریسی به صورت درایه به درایه انجام می‌شود:

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

◀ ضرب ماتریسی با ضرب سطرهای ماتریس اول در ستون‌های ماتریس دوم:

$$(AB)_{ij} = \sum_k a_{ik} b_{kj}$$

◀ برای حل سیستم‌های معادلات خطی می‌توان از این مفاهیم استفاده کرد.





## تبدیل‌های خطی

▶ تبدیل خطی مثل یک «ماشین تغییر شکل» است که روی بردارها عمل می‌کند.

▶ هر برداری که وارد آن شود، خروجی قابل پیش‌بینی و منظم است. هیچ پیچیدگی غیرخطی یا اعوجاجی ندارد.

▶ بردار ورودی چرخش داده می‌شود، کشیده یا فشرده می‌شود، بازتاب می‌شود.

▶ اگر دو بردار در ورودی هم‌راستا باشند، در خروجی هم هم‌راستا می‌مانند.

▶ خط‌ها، صفحه‌ها و نسبت‌ها حفظ می‌شوند.



## تبدیل‌های خطی

تبدیل خطی  $T$  باید دو شرط را رعایت کند:

جمع‌پذیری:

$$T(u + v) = T(u) + T(v)$$

ضرب‌پذیری با عدد اسکالر:

$$T(c.u) = c.T(u)$$

اگر بردارهای ورودی را ترکیب کنیم، تبدیل خطی همان ترکیب را روی خروجی‌ها اعمال می‌کند.

$$T(a.u + b.v) = a.T(u) + b.T(v)$$



## تبدیل‌های خطی

هر تبدیل خطی در فضای برداری می‌تواند به وسیله یک ماتریس نشان داده شود که بردار ورودی را به بردار خروجی نگاشت می‌کند:

$$y = Ax$$

حفظ جمع

$$A(u + v) = Au + Av$$

حفظ ضرب اسکالر

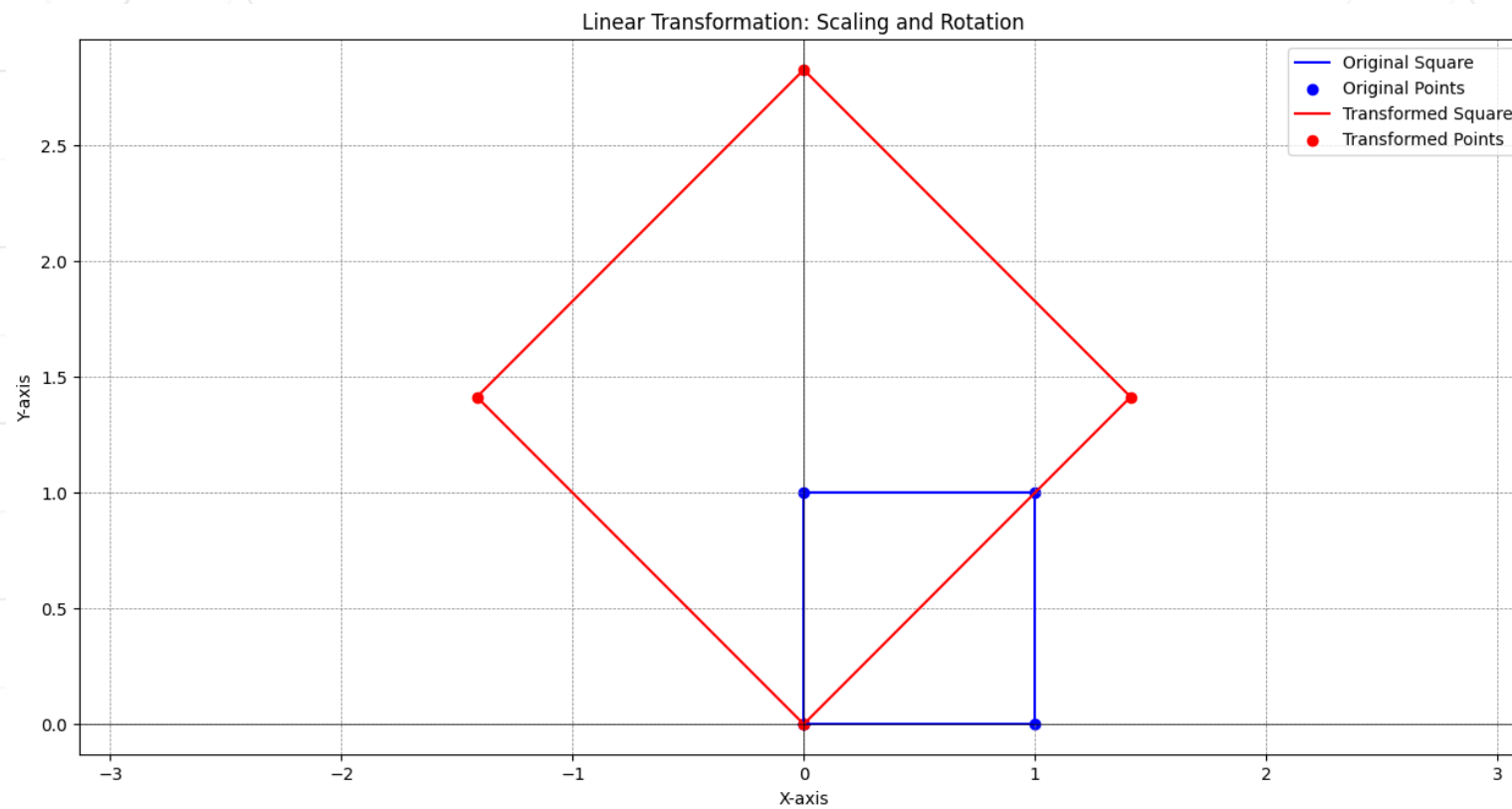
$$A(cv) = cAv$$

استفاده از ماتریس‌های چرخش برای تغییر جهت داده‌ها هنگام پیش‌پردازش.

انجام نرمال‌سازی یا استانداردسازی با ماتریس‌های مخصوص.



# تصویر تبدیل خطی





## رتبه ماتریس/تبدیل خطی

- ▶ رتبه مثل «میزان استقلال اطلاعات» در یک ماتریس یا تبدیل خطی است.
- ▶ در نظر بگیرید هر ستون ماتریس یک بردار در فضا است.
- ▶ اگر هر یک از بردارهای ستونی جهت متفاوتی داشته باشند و هیچ کدام ترکیب خطی دیگری نباشند، یعنی اطلاعات جدیدی دارند.
- ▶ اگر همه بردارها روی یک راستا/خط باشند → رتبه 1
- ▶ اگر روی یک صفحه باشند ولی نه روی یک خط → رتبه 2
- ▶ اگر فضا را پر کند (در فضای  $n$  بعدی) → رتبه  $n$
- ▶ رتبه می‌گویند «چند جهت مستقل» در ماتریس وجود دارد. عبارت دیگر، رتبه تبدیل خطی برابر است با بعد تصویر تبدیل.



## ماتریس رتبه کامل

▶ رتبه یک ماتریس به تعداد ستون‌های مستقل خطی یا سطرهای مستقل خطی آن اشاره دارد.

▶ یعنی رتبه، بعد فضای برداری‌ای است که توسط ستون‌ها یا سطرهای ماتریس ساخته می‌شود.

▶ رتبه تعداد بُعدها یا اطلاعات منحصر به فردی است که ماتریس دارد.

▶ یک ماتریس مربعی رتبه کامل، معکوس پذیر است



## جمع‌بندی مفهوم رتبه

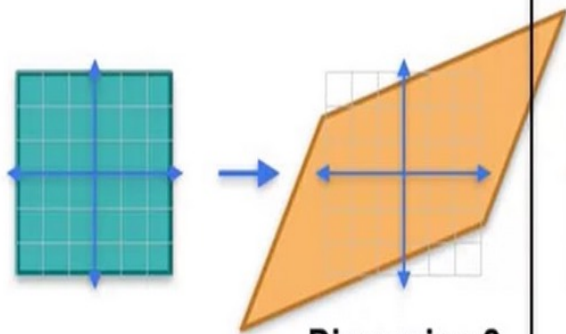
مفهوم	تعریف	تفسیر
رتبه ماتریس	تعداد ستون‌های مستقل خطی	چند جهت مستقل در ستون‌ها
رتبه تبدیل خطی	بعد تصویر تبدیل	تعداد ابعاد قابل دستیابی در فضای خروجی



# رتبه تبدیل خطی

Rank 2

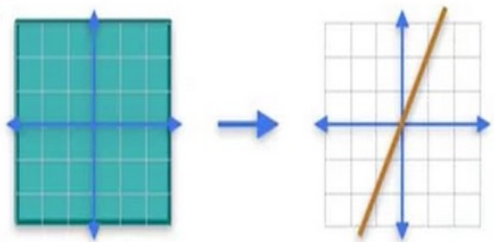
3	1
1	2



Dimension 2

Rank 1

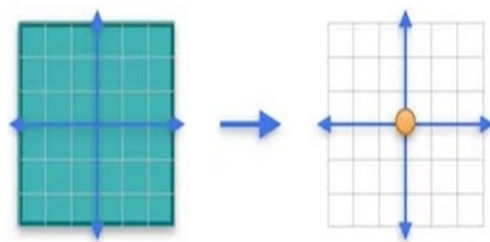
1	1
2	2



Dimension 1

Rank 0

0	0
0	0



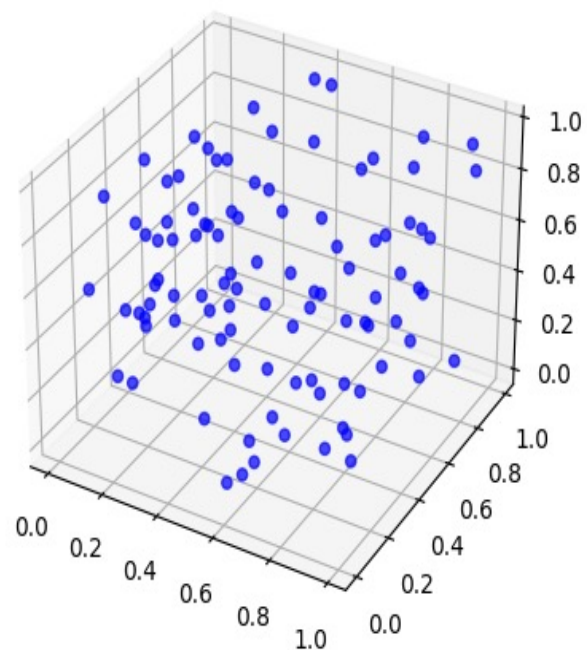
Dimension 0



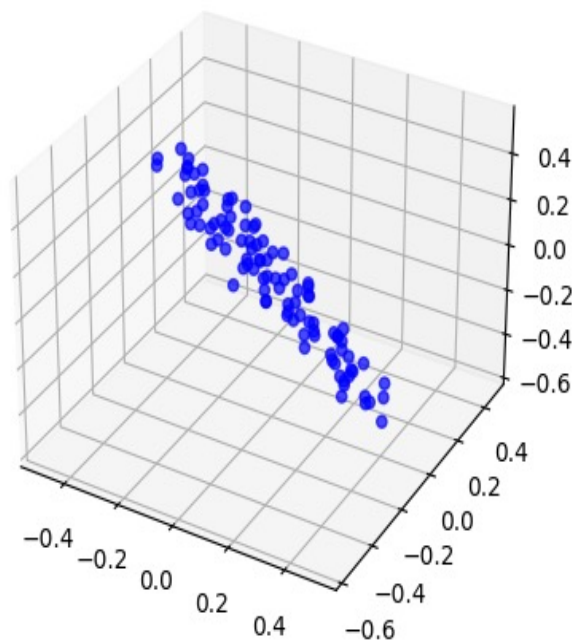


# خروجی نمونه کد تبدیل رتبه ناکامل

Original Data in 3D Space



Transformed Data in 3D Space



## دترمینان

◀ دترمینان عددی است که برخی خصوصیات یک ماتریس (مانند معکوس پذیری) را تعیین می‌کند.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \rightarrow |A| = a_{11}a_{22} - a_{12}a_{21}$$

◀ محاسبه دترمینان یک ماتریس داده جهت تصمیم‌گیری درباره‌ی استفاده از روش‌های حل معادلات خطی.



## معکوس ماتریس

▶ ماتریس  $A^{-1}$  معکوس ماتریس  $A$  است، به طوری که  $AA^{-1} = I$  که  $I$  ماتریس همانی است.

▶ در صورتی که دترمینان غیرصفر باشد، ماتریس معکوس پذیر است.

▶ برای حل دستگاه معادلات خطی  $b = Ax$  می توان از معکوس استفاده کرد.

$$x = A^{-1}b$$

▶ در کاربردهای رگرسیون ساده می توان به کمک معکوس ماتریس یک سیستم معادلات خطی را حل کرد.



## مثال: رگرسیون خطی

فرض می‌کنیم بین متغیرهای ورودی و خروجی یک سیستم ارتباطی از نوع رگرسیون خطی برقرار است:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_d x_d$$

دیتاستی با  $n$  داده با بعد  $d$  در اختیار داریم

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

می‌خواهیم پارامترهای مدل رگرسیون خطی (بردار  $\beta$ ) را تخمین بزنیم





## مثال: رگرسیون خطی

◀ اگر آنگاه  $b=0$  این تبدیل کاملاً خطی و معادل تبدیل خطی تعریف شده در بالا است.

◀ در غیر اینصورت، یک تبدیل affine است که در بسیاری از کاربردها هنوز هم به عنوان «تبدیل خطی» شناخته می‌شود:

◀ چرخش (rotation)، کشش یا فشردگی (scaling)، بازتاب (reflection)، جابجایی (translation)

◀ خطوط، همچنان خط باقی می‌مانند.

◀ نقاطی که روی یک خط بودند، بعد از تبدیل هم روی یک خط هستند.

◀ نسبت فاصله‌ها بین نقاط روی یک خط حفظ می‌شود.



## مثال: رگرسیون خطی

$$\bar{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nd} \end{bmatrix}$$

فرم ماتریسی

$$\bar{X} \beta = y$$

حل دستگاه معادلات خطی

$$\bar{X} \beta = y \Rightarrow \bar{X}^T \bar{X} \beta = \bar{X}^T y \Rightarrow \beta = (\bar{X}^T \bar{X})^{-1} \bar{X}^T y$$



## حل دستگاه معادلات خطی

$$\begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

$$\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

```
import numpy as np

# Define the coefficient matrix A
A = np.array([[4, 3],
              [2, 1]])

print(A)

# Define the constants vector B
b = np.array([20, 10])

# Calculate the determinant of matrix A
determinant = np.linalg.det(A)

# Check if the determinant is zero (singular matrix)
if determinant == 0:
    print("Matrix A is singular and cannot be inverted.")
else:
    # Calculate the inverse of matrix A
    A_inverse = np.linalg.inv(A)
    x = np.dot(A_inverse, b)
    print("\nSolution (X):")
    print("x0 =", x[0])
    print("x1 =", x[1])
```



## زیر فضا

◀ یک زیر فضا  $W$  از فضای برداری  $V$ ، زیرمجموعه‌ای از  $V$  است که خودش یک فضای برداری با عملیات تعریف شده در  $V$  محسوب می‌شود.

◀ وجود بردار صفر  $0 \in W$

◀ بسته بودن نسبت به جمع

$$u, v \in W \rightarrow u + v \in W$$

◀ بسته بودن نسبت به ضرب اسکالر

$$u \in W, c \in \mathbb{R} \rightarrow cu \in W$$





## زیر فضا: مثال‌ها

◀ در فضای  $\mathbb{R}^3$  یک خط یا یک صفحه که از مبدأ عبور می‌کند زیرفضا محسوب می‌شوند.

◀ با محاسبه مولفه‌های اصلی (Principal Components)، داده‌های با ابعاد بالا به یک فضای با بعد کمتر نگاشت می‌شوند.

◀ الگوریتم SVM سعی می‌کند یک ابرصفحه (یک زیرمجموعه از فضای برداری) پیدا کند که کلاس‌ها را از یکدیگر جدا کند.

◀ بسیاری از الگوریتم‌های خوشه‌بندی فرض می‌کنند که نقاط داده در چندین زیرفضای متفاوت قرار دارند.



## فضای تهی

◀ فضای تهی (یا هسته، Kernel) یک ماتریس  $A$  مجموعه‌ای از بردارهای  $x$

است که در آن‌ها رابطه  $Ax = 0$  برقرار است:

$$Null(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$$

◀ فضای تهی یک زیرفضا از فضای برداری اصلی است

◀ اگر ماتریس  $A$  دارای  $n$  ستون باشد، بعد فضای تهی عبارت است از:

$$nullity(A) = n - Rank(A)$$



## فضای تهی: مثال‌ها

▶ در مسائل رگرسیون خطی، ماتریس  $X$  شامل ویژگی‌های داده است. اگر ستون‌های  $X$  دچار هم‌بستگی باشند، برخی از آن‌ها به صورت خطی وابسته خواهند بود. این یعنی وجود بردارهای غیر صفر  $v$  در فضای تهی به گونه‌ای است که:

$$Xv = 0$$

▶ در تحلیل مولفه‌های اصلی (PCA)، ماتریس کوواریانس داده‌ها تشکیل می‌شود. جهت‌هایی که داده‌ها هیچ واریانسی در آن‌ها ندارند، متناظر با بردارهایی هستند که برابر با فضای تهی هستند.



## مقادیر ویژه و بردارهای ویژه

- ▶ در نظر بگیرید یک ماتریس مانند یک «تغییر شکل دهنده» در فضا عمل می‌کند.
- ▶ وقتی این ماتریس روی یک بردار عمل می‌کند، معمولاً جهت و اندازه‌ی آن بردار تغییر می‌کند.
- ▶ اما گاهی بردار (های) خاصی وجود دارد که وقتی ماتریس روی آن اعمال می‌شود، فقط طول آن تغییر می‌کند و جهت آن ثابت می‌ماند.
- ▶ این بردار خاص، بردار ویژه (Eigenvector) نامیده می‌شود و و ضربی که طول بردار را تغییر می‌دهد، مقدار ویژه (Eigenvalue) نامیده می‌شود.



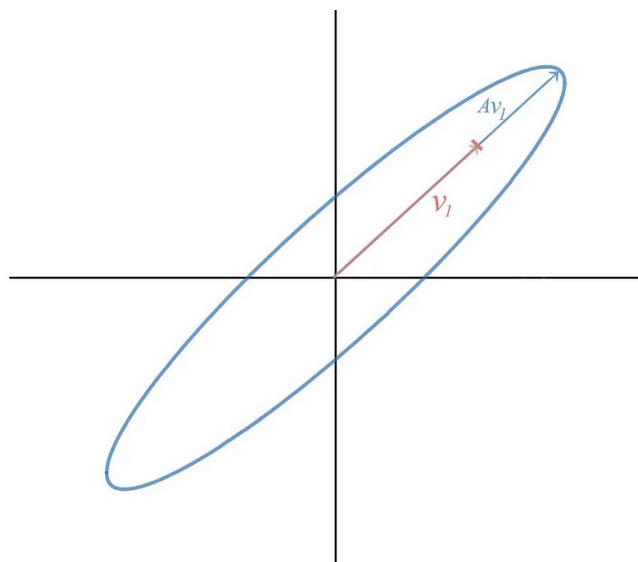


## مقادیر ویژه و بردارهای ویژه

برای ماتریس  $A$ ، اگر بردار  $v$  و اسکالر  $\lambda$  وجود داشته باشند که

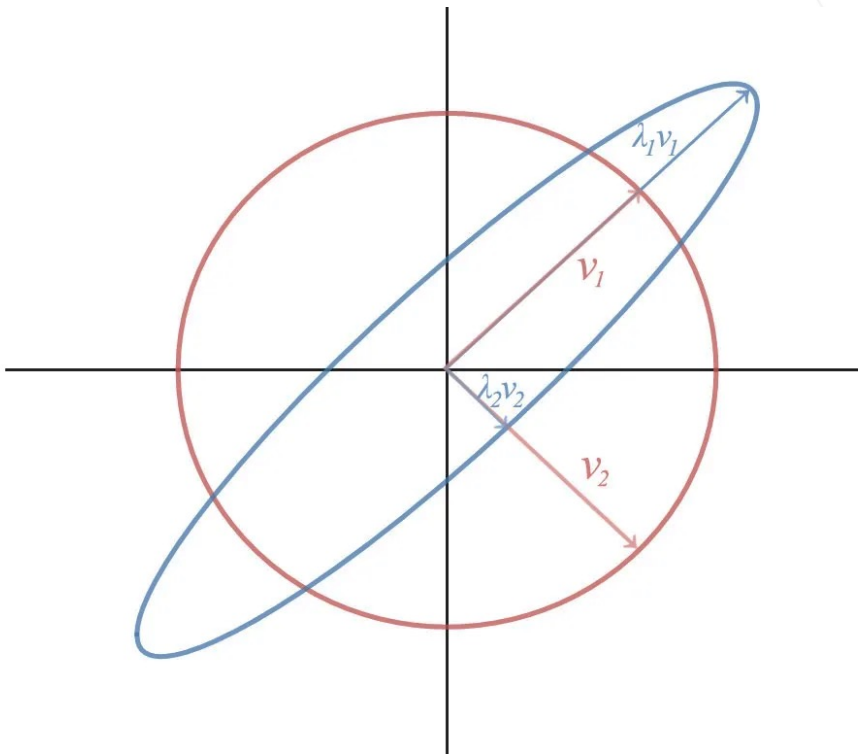
$$Av = \lambda v$$

آنگاه  $v$  بردار ویژه و  $\lambda$  مقدار ویژه نامیده می‌شود.



## مقادیر ویژه و بردارهای ویژه

◀ اگر مقدار ویژه بزرگتر از یک باشد،  $Av_i$  مربوطه منبسط می‌شود و بالعکس



## قطری سازی

◀ در نظر بگیرید  $A$  یک ماتریس  $2 \times 2$  است

$$A[v_1 \ v_2] = [\lambda_1 v_1 \ \lambda_2 v_2] = [v_1 \ v_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$$AV = V\Lambda$$

◀ ماتریس مربعی  $A$  قطری پذیر است اگر بتوانیم آن را به یک ماتریس

قطری تبدیل کنیم:

$$V^{-1}AV = \Lambda$$

◀ اگر ماتریس  $n \times n$  قطری پذیر باشد آنگاه  $n$  بردار ویژه مستقل خطی دارد.



## ویژگی‌های مهم مقادیر و بردارهای ویژه

◀ حاصلضرب مقادیر ویژه برابر با دترمینان است.

◀ برای محاسبه توان‌های ماتریس می‌توان از آن‌ها استفاده کرد:

$$A^k = V\Lambda^k V^{-1}$$

◀ اگر  $A$  منفرد باشد، مقدار ویژه 0 دارد.

◀ مقادیر ویژه و بردارهای ویژه می‌توانند اعداد مختلط باشند.





## بردارها و مقادیر ویژه چه می‌گویند؟

◀ وقتی ماتریس کوواریانس داده‌ها رو محاسبه می‌کنیم، در واقع داریم بررسی می‌کنیم که داده‌ها در کدام جهت‌ها بیشترین تغییر (واریانس) را دارند.

◀ با تجزیه این ماتریس به بردارهای ویژه و مقادیر ویژه، به این نتایج می‌رسیم:

◀ بردارهای ویژه: جهت‌هایی در فضای ویژگی‌ها که داده‌ها بیشترین یا کمترین تغییر را در آن‌ها دارند.

◀ مقادیر ویژه: میزان واریانس داده‌ها در امتداد هر بردار ویژه.



## ارتباط با کاهش بعد

- ◀ بعد واقعی داده‌ها (Intrinsic Dimensionality) یعنی تعداد جهت‌هایی که واقعاً داده‌ها در آن‌ها پخش شده‌اند.
- ◀ مثلاً اگر داده‌ای در فضای 100 بعدی باشد، ولی فقط در 3 جهت اصلی تغییر کند، بعد واقعی آن 3 است. در این حالت، فقط 3 مقدار ویژه بزرگ داریم و بقیه تقریباً صفرند.
- ◀ مقادیر ویژه بزرگ‌تر → جهت‌هایی که اطلاعات بیشتری دارند
- ◀ مقادیر ویژه کوچک‌تر → جهت‌هایی که تقریباً نویز هستند یا اطلاعات کمی دارند
- ◀ با انتخاب بردارهای ویژه متناظر با بزرگ‌ترین مقادیر ویژه، می‌توانیم داده‌ها را روی یک فضای کوچکتر فشرده کنیم، بدون اینکه بخش مهمی از اطلاعات را از دست بدهیم.



## تجزیه مقادیر منفرد

◀ SVD ابزاری قدرتمند برای تجزیه و تحلیل داده‌ها است که در تحلیل‌های یادگیری ماشین نقش اساسی دارد.

◀ تعمیمی از تجزیه مقادیر ویژه است

◀ برای یک ماتریس  $A$  با ابعاد  $m \times n$ :

$$A = U\Sigma V^T$$

◀ ماتریس  $U$  با ابعاد  $m \times m$  و ستون‌های متعامد

◀ ماتریس قطری  $\Sigma$  با عناصر غیر صفر مرتب نزولی به عنوان مقادیر منفرد

◀ ماتریس  $V^T$  ترانزاده ماتریس متعامد  $V$  با ابعاد  $n \times n$



## خصوصیات و نکات کلیدی

### ◀ ارتباط با مقادیر ویژه

◀ SVD به طور غیر مستقیم به تحلیل مقادیر ویژه  $AA^T$  مرتبط است.

◀ مقادیر منفرد  $\sigma_i$  ریشه های مثبت مقادیر ویژه  $AA^T$  هستند.

### ◀ کاهش بعد

◀ در الگوریتم هایی نظیر PCA، استفاده از SVD به جداسازی مولفه های اصلی داده ها و بهبود عملکرد مدل های یادگیری ماشین کمک می کند.

### ◀ فشرده سازی تصاویر و توصیه گرها

◀ با نگه داشتن تنها مقادیر منفرد بزرگ، داده های اصلی با نویز و اطلاعات جزئی حذف شده به دست می آید.

### ◀ کاهش نویز در داده ها

◀ نگه داشتن تنها مؤلفه های اصلی و حذف نویزهای کوچک.





## سوالات رایج

تفاوت بین ضرب داخلی و ضرب ماتریسی چیست؟

چگونه می‌توان از مفهوم مقادیر ویژه در کاهش ابعاد استفاده کرد؟

چه ارتباطی بین PCA و SVD وجود دارد؟

موارد ذیل را برای بردار ویژه و مقدار ویژه ثابت کنید:

$$A^k v = \lambda^k v$$

$$A^{-1} v = \lambda^{-1} v$$

$$(A + cI)v = (\lambda + c)v$$



## منابع پیشنهادی

- ▶ Strang, G. (2000). Linear algebra and its applications.
- ▶ Golub, G. H., & Van Loan, C. F. (2013). Matrix computations. JHU press.

