

The Singular Value Decomposition

This is the capstone of the class.

It's a fundamental application of linear algebra to:

- Statistics (PCA)
- Engineering
- Data Science
- etc.

Today we'll discuss the **outer product form** and the **mechanics** (plumbing?) of the SVD.

Introduction to the SVD

Thm (SVD, outer product form):

back to
rectangular
matrices

Let A be an $m \times n$ matrix of rank r . Then

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T$$

where

- $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$
- $\{u_1, \dots, u_r\}$ is an **orthonormal** set in $\mathbb{R}^m$
- $\{v_1, \dots, v_r\}$ is an **orthonormal** set in $\mathbb{R}^n$.

What does this mean?

Idea: columns of A are data points

Here's an informal description of what SVD says.

$r=1$:

let $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$ be nonzero vectors.

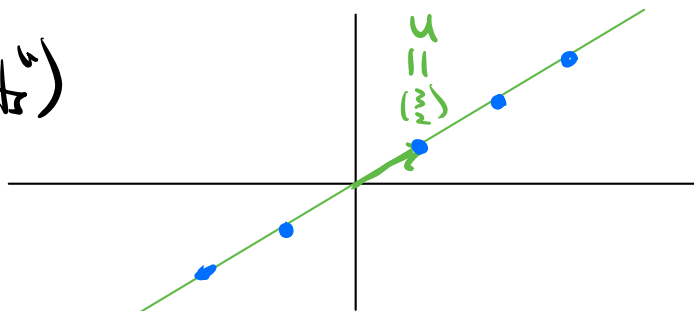
$$uv^T = \underbrace{\begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}}_{\text{vector}} \underbrace{(v_1 \dots v_n)}_{\text{weights}} = \begin{pmatrix} v_1 u & \dots & v_n u \end{pmatrix}$$

multiples of u

This is an $m \times n$ matrix of rank 1: $\text{Col}(uv^T) = \text{Span}\{u\}$

Let's plot the columns ("data points")

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



Upshot: A rank-1 matrix encodes data points (columns) that lie on a line ($\dim \text{Col}(A) = 1$). The SVD tells you which line & which multiples.

$r=2$:

$$A = \underbrace{u_1}_{\text{vector}} \underbrace{v_1^T}_{\text{weights}} + \underbrace{u_2}_{\text{vector}} \underbrace{v_2^T}_{\text{weights}} = \begin{pmatrix} v_{11}u_1 & \dots & v_{1n}u_1 \end{pmatrix} + \begin{pmatrix} v_{21}u_2 & \dots & v_{2n}u_2 \end{pmatrix}$$
$$= \begin{pmatrix} v_{11}u_1 + v_{21}u_2 & \dots & v_{1n}u_1 + v_{2n}u_2 \end{pmatrix}$$

The columns are linear combinations of u_1 & u_2 .

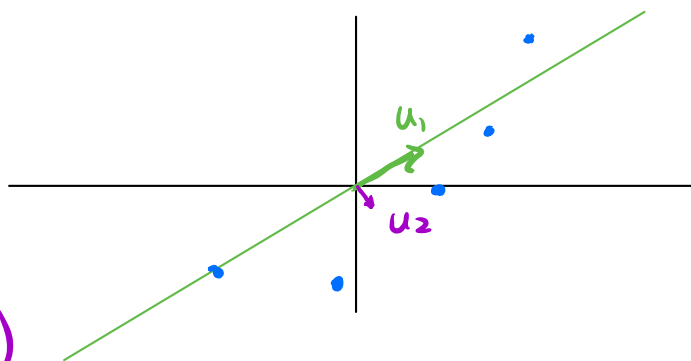
Let's plot the columns ("data points"):

$$u_1 = \text{weights of } \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

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

$$+ \underbrace{\begin{pmatrix} 2 \\ -3 \end{pmatrix}}_{\text{orthogonal}} \begin{pmatrix} 3 & 1 & 2 & -1 & 0 \end{pmatrix}$$

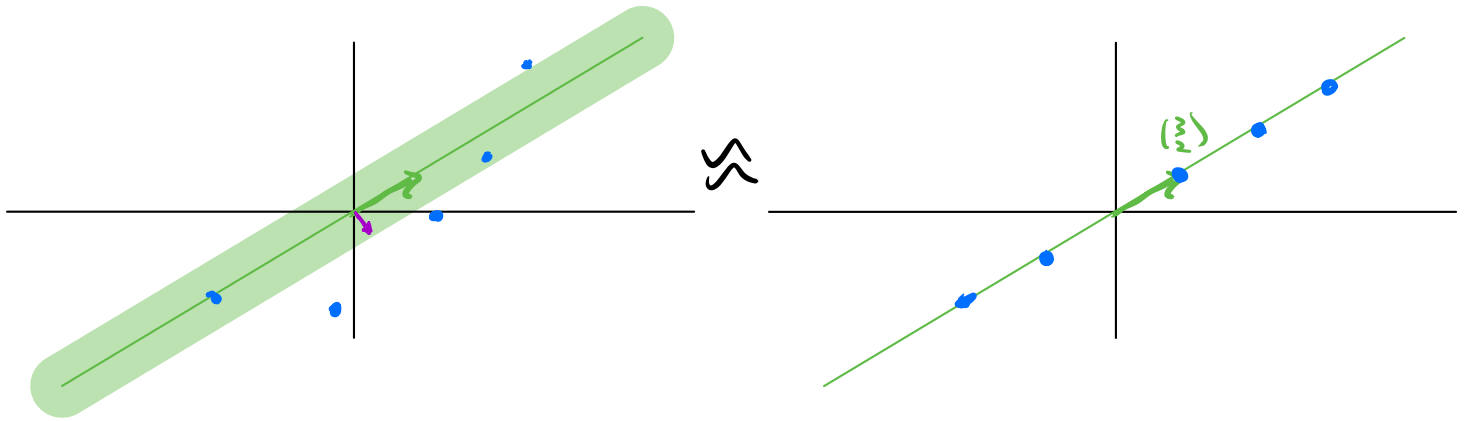
$$u_2 \quad v_2 = \text{weights of } \begin{pmatrix} 2 \\ -3 \end{pmatrix}$$



Upshot: A rank-2 matrix encodes **data points** that lie on a **plane** ($\dim \text{Col}(A) = 2$). The SVD gives you a **basis** $\{u_1, u_2\}$ and the **weights** for each column.

But: $\| \begin{pmatrix} 3 \\ 2 \end{pmatrix} \| \gg \| \begin{pmatrix} .2 \\ -.3 \end{pmatrix} \|$ so the $\begin{pmatrix} .2 \\ -.3 \end{pmatrix}$ direction is **less important!**

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix} (-1 \ 2 \ 1 \ 3 \ -2) + \begin{pmatrix} .2 \\ -.3 \end{pmatrix} (3 \ 1 \ 2 \ -1 \ 0) \\ \approx \begin{pmatrix} 3 \\ 2 \end{pmatrix} (-1 \ 2 \ 1 \ 3 \ -2) \quad (\text{to one decimal place})$$



We've extracted important information:
our data points **almost lie on a line!**

In general, the SVD will find the **best-fit** line, plane, 3-space, ..., r -space for our data, all at once, and tell you **how good is the fit** in the sense of **orthogonal least squares!**

(more on this later)

Why might we care?

- **Data compression:** uv^T is 7 numbers instead of 10 for a 2×5 matrix.
- **Data analysis:** SVD will reveal all approximate linear relations among our data points.
- **Dimension reduction:** if our data in $\mathbb{R}^{100000}$ almost lie on a 100-dimensional subspace then computers are happier to do the computations.
- **Statistics:** SVD finds more & less important correlations etc.

Mechanics of the SVD

Back to the statement of the SVD:

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T \quad r = \text{rank}(A)$$

where

- $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$
- $\{u_1, \dots, u_r\}$ is an orthonormal set in $\mathbb{R}^m$
- $\{v_1, \dots, v_r\}$ is an orthonormal set in $\mathbb{R}^n$.

Def:

- $\sigma_1, \dots, \sigma_r$ are the singular values of A
- $u_1, \dots, u_r$ are the left singular vectors
- $v_1, \dots, v_r$ are the right singular vectors

Here are some **formal consequences** of the statement.

Note 1: For any vector $x \in \mathbb{R}^n$,

$$Ax = (\sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T)x = \sigma_1 u_1 v_1^T x + \dots + \sigma_r u_r v_r^T x \\ = \sigma_1 (v_1 \cdot x) u_1 + \dots + \sigma_r (v_r \cdot x) u_r$$

$$Ax = \sigma_1 (v_1 \cdot x) u_1 + \dots + \sigma_r (v_r \cdot x) u_r$$

Note 2: Taking $x = v_i$, we have

$$Av_i = \sigma_1 (v_1 \cdot v_i) u_1 + \dots + \sigma_i (v_i \cdot v_i) u_i + \dots + \sigma_r (v_r \cdot v_i) u_r$$

(recall $\{u_1, \dots, u_r\}$ and $\{v_1, \dots, v_r\}$ are **orthonormal**).

So the singular vectors are related by

$$Av_i = \sigma_i u_i$$

and thus

$$\|Av_i\| = \sigma_i$$

Note 3: Take transposes:

$$A^T = (\sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T)^T = \sigma_1 v_1 u_1^T + \dots + \sigma_r v_r u_r^T$$

Therefore, $A^T = \sigma_1 v_1 u_1^T + \dots + \sigma_r v_r u_r^T$

is the **SVD** of A^T !

So A & A^T have the same

- **singular values** and
- **singular vectors** (switch right & left).

Note 4: Note 2 + Note 3 $\Rightarrow A^T u_i = \sigma_i v_i$, so

$$A^T A v_i = A^T (\sigma_i u_i) = \sigma_i A^T u_i = \sigma_i^2 v_i$$

$$A A^T u_i = A (\sigma_i v_i) = \sigma_i A v_i = \sigma_i^2 u_i$$

$$A^T A v_i = \sigma_i^2 v_i$$

$$A A^T u_i = \sigma_i^2 u_i$$

In particular,

$\{v_1, \dots, v_r\}$ are orthonormal eigenvectors of $A^T A$
with eigenvalues $\sigma_1^2, \dots, \sigma_r^2$.

$\{u_1, \dots, u_r\}$ are orthonormal eigenvectors of $A A^T$
with eigenvalues $\sigma_1^2, \dots, \sigma_r^2$.

This tells us how to prove/compute the SVD:
orthogonally diagonalize $A^T A$

Proof of the SVD: pay attention to steps 1-3: they illustrate the mechanics of the SVD!

Let $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ be the eigenvalues of $A^T A$
(the λ_i 's show up multiple times if $A/M \geq 1$)

Note $\lambda_n \geq 0$ because $A^T A$ is positive-semidefinite.

Step 1: I claim $\lambda_{r+1} = \dots = \lambda_n = 0$.

- $\text{Nul}(A^T A) = \text{Nul}(A)$ has dimension $n-r$.
- $\text{Nul}(A^T A) =$ the 0-eigenspace of $A^T A$.

- $AM(0) = GM(0)$ in $A^T A$

because $A^T A$ is symmetric $\Rightarrow$ diagonalizable

So $n-r$ of the λ_i 's are $= 0$

$$\Rightarrow \lambda_{r+1} = \dots = \lambda_n = 0$$

Now: $\lambda_1 \geq \dots \geq \lambda_r > 0$ are the
nonzero eigenvalues of $A^T A$.

Set:

- $\sigma_i = \sqrt{\lambda_1}, \dots, \sigma_r = \sqrt{\lambda_r}$

- Let $v_1, \dots, v_r$ be orthonormal eigenvectors
with $A^T A v_i = \lambda_i v_i$.

- Let $u_i = \frac{1}{\sigma_i} A v_i, \dots, u_r = \frac{1}{\sigma_r} A v_r$

Step 2: I claim $\{u_1, \dots, u_r\}$ is orthonormal. Check:

$$\begin{aligned} u_i \cdot u_j &= u_i^T u_j = \left(\frac{1}{\sigma_i} A v_i \right)^T \left(\frac{1}{\sigma_j} A v_j \right) = \frac{1}{\sigma_i \sigma_j} (A v_i)^T (A v_j) \\ &= \frac{1}{\sigma_i \sigma_j} (v_i^T A^T A v_j) = \frac{1}{\sigma_i \sigma_j} v_i^T (A^T A v_j) = \frac{1}{\sigma_i \sigma_j} v_i^T (\lambda_j v_j) \\ &= \frac{\sigma_j^2}{\sigma_i \sigma_j} v_i^T v_j = \frac{\sigma_j}{\sigma_i} v_i \cdot v_j \end{aligned}$$

Since $\{v_1, \dots, v_r\}$ is orthonormal:

- If $i=j$ this is $u_i \cdot u_i = \frac{\sigma_i}{\sigma_i} v_i \cdot v_i = \|v_i\|^2 = 1$

- If $i \neq j$ this is $\frac{\sigma_j}{\sigma_i} v_i \cdot v_j = 0$

Step 3: I claim $\{v_1, \dots, v_r\}$ is a basis for $\text{Row}(A)$.

- $v_i = \frac{1}{\lambda_i} A^T A v_i = A^T \left(\frac{1}{\lambda_i} A v_i \right) \in \text{Col}(A^T) = \text{Row}(A)$
- $\dim \text{Row}(A) = r$ and $\{v_1, \dots, v_r\}$ is orthonormal
 $\Rightarrow$ linearly independent

So the Basis Theorem $\Rightarrow \text{Row}(A) = \text{Span}\{v_1, \dots, v_r\}$

Step 4: Verify $A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T$.

Let $B = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T$, so we want to show $A \stackrel{?}{=} B$.

Recall $A = B$ if $Ax = Bx$ for all $x \in \mathbb{R}^n$.

As above,

$$Bx = \sigma_1 (v_1 \cdot x) u_1 + \dots + \sigma_r (v_r \cdot x) u_r.$$

$$\begin{aligned} Bv_i &= \sigma_1 (v_1 \cdot v_i) u_1 + \dots + \sigma_i (v_i \cdot v_i) u_i + \dots + \sigma_r (v_r \cdot v_i) u_r \\ &= \sigma_i u_i \end{aligned}$$

(1) If $x \in \text{Nul}(A)$ then $Ax = 0$ and

$$Bx = \sigma_1 (v_1 \cdot x) u_1 + \dots + \sigma_r (v_r \cdot x) u_r = 0 = Ax$$

because $v_1, \dots, v_r \in \text{Row}(A) = \text{Nul}(A)^\perp$ ✓

(2) If $x \in \text{Row}(A)$ then we can solve

$$x = x_1 v_1 + \dots + x_r v_r \quad \text{by Step 3. Then}$$

$$Ax = A(x_1 v_1 + \dots + x_r v_r) = x_1 A v_1 + \dots + x_r A v_r$$

$$(u_i = \frac{1}{\sigma_i} A v_i) = x_1 \sigma_1 u_1 + \dots + x_r \sigma_r u_r$$

$$Bx = B(x_1v_1 + \dots + x_rv_r) = x_1Bv_1 + \dots + x_rBv_r$$

$$(Bv_i = \sigma_i u_i) = x_1\sigma_1 u_1 + \dots + x_r\sigma_r u_r = Ax \quad \checkmark$$

(3) Any $x \in \mathbb{R}^n$ has an orthogonal decomposition
 $x = x_v + x_{v^\perp}$ $x_v \in \text{Row}(A)$ $x_{v^\perp} \in \text{Nul}(A)$.

$$\Rightarrow Ax = A(x_v + x_{v^\perp}) = Ax_v + Ax_{v^\perp}$$

$$\stackrel{(1,2)}{=} Bx_v + Bx_{v^\perp} = B(x_v + x_{v^\perp}) = Bx \quad \checkmark$$

NB: AA^T and $A^T A$ have the same nonzero eigenvalues $\sigma_1^2, \dots, \sigma_r^2$. (We showed in the proof that the other eigenvalues are $= 0$.)

→ What about the 0 eigenvalue?

→ What if A is a tall matrix with FCR?

NB: We showed in the proof that

$\{v_1, \dots, v_r\}$ is a basis for $\text{Row}(A)$.

Replace A by $A^T \rightsquigarrow$

$\{u_1, \dots, u_r\}$ is a basis for $\text{Row}(A^T) = \text{Col}(A)$.

Mechanics of the SVD: Summary

A : an $m \times n$ matrix of rank r

SVD: $A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T$

$$Ax = \sigma_1 (v_1 \cdot x) u_1 + \dots + \sigma_r (v_r \cdot x) u_r$$

$\sigma_1 \geq \dots \geq \sigma_r > 0$: singular values

$\sigma_i^2 \geq \dots \geq \sigma_r^2$: nonzero eigenvalues of $A^T A$ and AA^T

$\{u_1, \dots, u_r\}$:

- left singular vectors
- orthonormal eigenvectors of AA^T

$$AA^T u_i = \sigma_i^2 u_i$$

• orthonormal basis for $\text{Col}(A)$

$\{v_1, \dots, v_r\}$:

- right singular vectors
- orthonormal eigenvectors of $A^T A$

$$A^T A v_i = \sigma_i^2 v_i$$

• orthonormal basis for $\text{Row}(A)$

$$A v_i = \sigma_i u_i \Rightarrow \|A v_i\| = \sigma_i$$

SVD: $A^T = \sigma_1 v_1 u_1^T + \dots + \sigma_r v_r u_r^T$

$$A^T u_i = \sigma_i v_i \Rightarrow \|A^T u_i\| = \sigma_i$$

This also gives us a procedure to compute the SVD.
It is **not** the algorithm used in practice!

→ Efficient computation of the SVD is a difficult problem!

Naive Schoolbook Procedure to Compute the SVD:

Let A be an $m \times n$ matrix of rank r .

(1) Compute the nonzero eigenvalues of $A^T A$,
 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$

(where λ_i appears multiple times if $\text{AM} > 1$)

→ There are automatically r of them, and they're positive.

(2) Find an orthonormal eigenbasis for each eigenspace: get an orthonormal set $\{v_1, \dots, v_r\}$ with $A^T A v_i = \lambda_i v_i$.

(3) Set $\sigma_i = \sqrt{\lambda_i}$, $u_i = \frac{1}{\sigma_i} A v_i$.

Then $\{u_1, \dots, u_r\}$ is orthonormal and

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T.$$

NB: It may be easier to compute SVD of A^T !

(if A is **wide**: $m < n$, $A^T A$ is $n \times n$
but $A A^T$ is $m \times m$)

Eg: $A = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}$ NB: $r=2$ (2 pivots)

$$(1) \quad A^T A = \begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix} \quad p(\lambda) = \lambda^2 - 50\lambda + 225 \\ = (\lambda - 45)(\lambda - 5)$$

$$\lambda_1 = 45 \quad \lambda_2 = 5$$

(2) Compute eigenspaces:

$$A^T A - 45 I_2 = \begin{pmatrix} -20 & 20 \\ - & - \end{pmatrix} \xrightarrow{\text{trick}} \begin{pmatrix} -b \\ a-\lambda \end{pmatrix} \quad v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$A^T A - 5 I_2 = \begin{pmatrix} 20 & 20 \\ - & - \end{pmatrix} \rightarrow v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$(3) \quad \sigma_1 = \sqrt{\lambda_1} = \sqrt{45} = 3\sqrt{5} \quad \sigma_2 = \sqrt{\lambda_2} = \sqrt{5}$$

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{3\sqrt{5}} \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{3\sqrt{10}} \begin{pmatrix} 3 \\ 9 \end{pmatrix} = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

SVD:

$$\begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} = 3\sqrt{5} \cdot \frac{1}{\sqrt{10}} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \end{pmatrix} + \sqrt{5} \cdot \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ -1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \end{pmatrix}$$

Check: $\|u_1\| = \frac{1}{\sqrt{10}} \sqrt{1^2 + 3^2} = 1 \quad \|u_2\| = \frac{1}{\sqrt{10}} \sqrt{3^2 + (-1)^2} = 1$

$$u_1 \cdot u_2 = 0 \quad \checkmark$$