The Singular Value Decomposition

This is the capstone of the class.
It's a fundamental application of linear algebra to:

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

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

Introduction to the SVD

back to rectangular matrices

Thm (SVD, outer product form):
Let A be an $m\times n$ matrix of rank r. Then

$$\boxed{A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T}$$

where

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

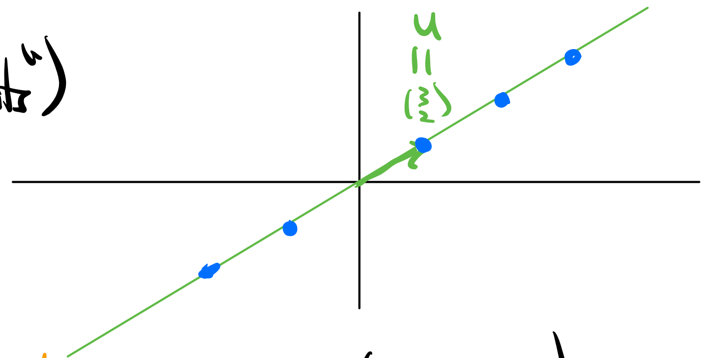
r=1: Let $u\in\mathbb{R}^m$, $v\in\mathbb{R}^n$ be nonzero vectors.

$$uv^T = \underset{\text{vector}}{\begin{pmatrix}u_1\\ \vdots\\ u_m\end{pmatrix}}\underset{\text{weights}}{(v_1\cdots v_n)} = \underset{\text{multiples of } u}{\begin{pmatrix} | & & | \\ v_1u & \cdots & v_nu \\ | & & |\end{pmatrix}}$$

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

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

$$\begin{pmatrix}3\\2\end{pmatrix}(-1 \quad 2 \quad 1 \quad 3 \quad -2)$$



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

r=2:

$$A = \underset{\text{weights}}{\overset{\text{vectors}}{u_1v_1^T + u_2v_2^T}} = \begin{pmatrix} | & & | \\ v_{11}u_1 & \cdots & v_{1n}u_1 \\ | & & |\end{pmatrix} + \begin{pmatrix} | & & | \\ v_{21}u_2 & \cdots & v_{2n}u_2 \\ | & & |\end{pmatrix}$$

$$= \overset{\text{weights}}{\begin{pmatrix} | & & | \\ v_{11}u_1 + v_{21}u_2 & \cdots & v_{1n}u_1 + v_{2n}u_2 \\ | & & |\end{pmatrix}}$$

The columns are linear combinations of u_1 & u_2.

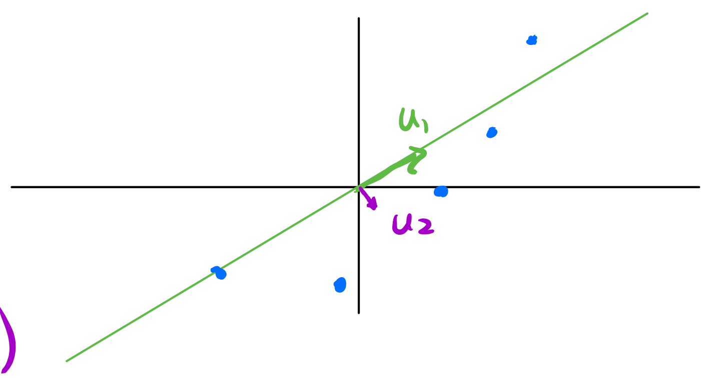
Let's plot the columns ("data points"):

v_1 = weights of $\begin{pmatrix}3\\2\end{pmatrix}$

$$\underset{u_1}{\begin{pmatrix}3\\2\end{pmatrix}}(-1 \quad 2 \quad 1 \quad 3 \quad -2) + \underset{u_2}{\begin{pmatrix}.2\\-.3\end{pmatrix}}(3 \quad 1 \quad 2 \quad -1 \quad 0)$$

orthogonal

v_2 = weights of $\begin{pmatrix}.2\\-.3\end{pmatrix}$

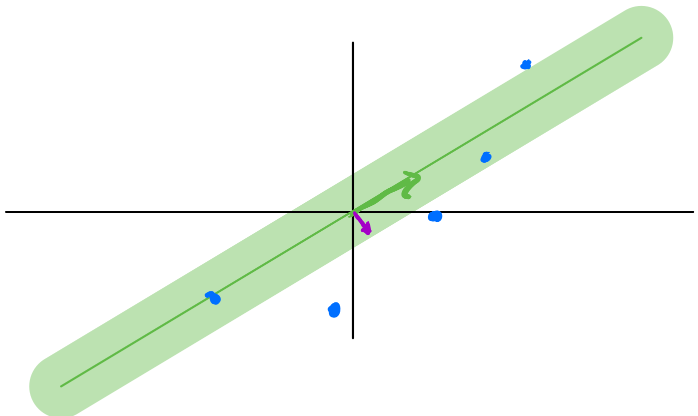


Upshot: A rank-2 matrix encodes data points that lie on a plane ($\dim \mathrm{Col}(A) = 2$). The SVD gives you a basis $\{u_1, u_2\}$ and the weights for each column.

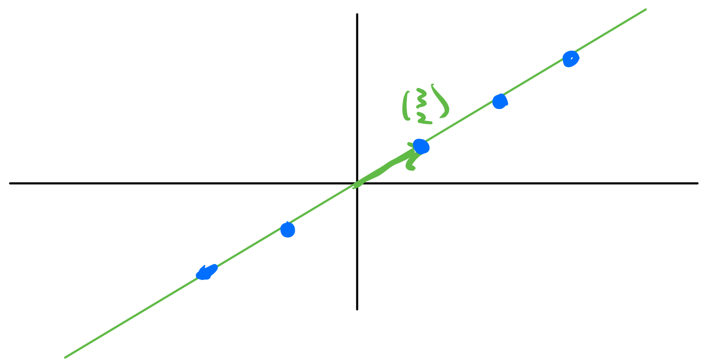
But: $\left\|\begin{pmatrix}3\\2\end{pmatrix}\right\| \gg \left\|\begin{pmatrix}.2\\-.3\end{pmatrix}\right\|$ so the $\begin{pmatrix}.2\\-.3\end{pmatrix}$ direction is less important!

$$\begin{pmatrix}3\\2\end{pmatrix}\begin{pmatrix}-1 & 2 & 1 & 3 & -2\end{pmatrix} + \begin{pmatrix}.2\\-.3\end{pmatrix}\begin{pmatrix}3 & 1 & 2 & -1 & 0\end{pmatrix}$$

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



$\approx$



We're extracted important information:

our data points almost lie on a line!

In general, the SVD will find the best-fit line, plane, 3-space, ..., n-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 $\mathbb{R}^{1000000}$ almost lie on a 1000-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:

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T \qquad r = \operatorname{rank}(A)$$

where

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

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

Here are some formal consequences of the statement.

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

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

$$\boxed{Ax = \sigma_1 (v_1 \cdot x) u_1 + \cdots + \sigma_r (v_r \cdot x) u_r}$$

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

$$Av_i = \sigma_1 \overset{=0}{(v_1 \cdot v_i)} u_1 + \cdots + \sigma_i \overset{=1}{(v_i \cdot v_i)} u_i + \cdots + \sigma_r \overset{=0}{(v_r \cdot v_i)} u_r$$

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

So the singular vectors are related by

$$\boxed{Av_i = \sigma_i u_i} \quad \text{and thus} \quad \boxed{\|Av_i\| = \sigma_i}$$

Note 3: Take transposes:

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

Therefore, $\boxed{A^T = \sigma_1 v_1 u_1^T + \cdots + \sigma_r v_r u_r^T}$

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

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

$$\boxed{\begin{aligned} A^TAv_i &= \sigma_i^2 v_i \\ AA^Tu_i &= \sigma_i^2 u_i \end{aligned}}$$

In particular

$$\boxed{\begin{aligned} &\{v_1, \ldots, v_r\} \text{ are orthonormal eigenvectors of } A^TA \\ &\qquad \text{with eigenvalues } \sigma_1^2, \ldots, \sigma_r^2. \\ &\{u_1, \ldots, u_r\} \text{ are orthonormal eigenvectors of } AA^T \\ &\qquad \text{with eigenvalues } \sigma_1^2, \ldots, \sigma_r^2. \end{aligned}}$$

This tells us how to prove/compute the SVD:

orthogonally diagonalize A^TA

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

Let $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of A^TA

(the λ_i's show up multiple times if $AM \geq 1$)

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

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

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

- $AM(0) = GM(0)$ in A^TA
  because A^TA is symmetric $\Rightarrow$ diagonalizable

So $n-r$ of the λ_i's are $=0$

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

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

Set:

- $\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_r = \sqrt{\lambda_r}$
- Let $v_1, \ldots, v_r$ be orthonormal eigenvectors with $A^TAv_i = \lambda_i v_i$.
- Let $u_1 = \frac{1}{\sigma_1}Av_1, \ldots, u_r = \frac{1}{\sigma_r}Av_r$

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

$$u_i \cdot u_j = u_i^T u_j = \left(\frac{1}{\sigma_i}Av_i\right)^T\left(\frac{1}{\sigma_j}Av_j\right) = \frac{1}{\sigma_i\sigma_j}(Av_i)^T(Av_j)$$

$$= \frac{1}{\sigma_i\sigma_j}(v_i^TA^TAv_j) = \frac{1}{\sigma_i\sigma_j}v_i^T(A^TAv_j) = \frac{1}{\sigma_i\sigma_j}v_i^T(\lambda_j v_j)$$

$$= \frac{\sigma_j^2}{\sigma_i\sigma_j}v_i^Tv_j = \frac{\sigma_j}{\sigma_i}v_i \cdot v_j$$

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

- If $i=j$ this is $u_i \cdot u_i = \overbrace{\frac{\sigma_i}{\sigma_i}}^{=1} v_i \cdot v_i = \|v_i\|^2 = 1$
- 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 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 $A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$.

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

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

As above,

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

$$Bv_i = \sigma_1 \underbrace{(v_1 \cdot v_i)}_{=0} u_1 + \cdots + \sigma_i \underbrace{(v_i \cdot v_i)}_{=1} u_i + \cdots + \sigma_r \underbrace{(v_r \cdot v_i)}_{=0} u_i$$
$$= \sigma_i u_i$$

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

$$Bx = \sigma_1 \underbrace{(v_1 \cdot x)}_{=0} u_1 + \cdots + \sigma_r \underbrace{(v_r \cdot x)}_{=0} u_r = 0 = Ax$$

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

$x = x_1 v_1 + \cdots + x_r v_r$ by Step 3. Then

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

$$\left(u_i = \tfrac{1}{\sigma_i} A v_i\right) = x_1 \sigma_1 u_1 + \cdots + x_r \sigma_r u_r$$

$$Bx = B(x_1 v_1 + \cdots + x_r v_r) = x_1 Bv_1 + \cdots + x_r Bv_r$$

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

(3) Any $x \in \mathbb{R}^n$ has an orthogonal decomposition

$$x = x_V + x_{V^\perp} \qquad x_V \in \mathrm{Row}(A) \quad x_{V^\perp} \in \mathrm{Nul}(A).$$

$$\Rightarrow Ax = A(x_V + x_{V^\perp}) = Ax_V + Ax_{V^\perp}$$

$$\overset{(1,2)}{=} Bx_V + Bx_{V^\perp} = B(x_V + x_{V^\perp}) = Bx \quad \checkmark$$

NB: A^TA and AA^T have the same nonzero eigenvalues $\sigma_1^2, \ldots, \sigma_r^2$. (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

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

Replace A by A^T ⇝

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

Mechanics of the SVD: Summary

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

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

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

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

$\sigma_1^2 \geq \cdots \geq \sigma_r^2$: nonzero eigenvalues of A^TA and AA^T

$\{u_1, \ldots, u_r\}$:
- left singular vectors
- orthonormal eigenvectors of AA^T

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

- orthonormal basis for $\mathrm{Col}(A)$

$\{v_1, \ldots, v_r\}$:
- right singular vectors
- orthonormal eigenvectors of A^TA

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

- orthonormal basis for $\mathrm{Row}(A)$

$$Av_i = \sigma_i u_i \implies \|Av_i\| = \sigma_i$$

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

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

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^TA,

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$$

(where λ_i appears multiple times if $AM > 1$)

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

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

(3) Set $\sigma_i = \sqrt{\lambda_i}$ $\quad u_i = \frac{1}{\sigma_i} Av_i$.

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

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

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

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

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

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

$\lambda_1 = 45 \qquad \lambda_2 = 5$

(2) Compute eigenspaces: $\begin{pmatrix} -b \\ a-\lambda \end{pmatrix}$

$A^TA - 45I_2 = \begin{pmatrix} -20 & 20 \\ \_ & \_ \end{pmatrix} \overset{\text{trick}}{\rightsquigarrow} v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

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

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

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

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

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

$u_1 \cdot u_2 = 0$ ✓