The Singular Value Decomposition: Introduction

L24

We finally come to the capstone of the class.
The SVD is a fundamental application of linear algebra to:

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

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

Theorem (SVD; outer product form):

(back to rectangular matrices)

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:

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$
- $\{u_1, u_2, \ldots, u_r\}$ is an orthonormal set in $\mathbb{R}^m$.
- $\{v_1, v_2, \ldots, v_r\}$ is an orthonormal set in $\mathbb{R}^n$.

What does this mean?

Idea: Think of the columns of A as data points.

Here's an informal description of what the SVD says. Let's not worry about the σ_i's or unit vectors yet.

$r=1$: If $u\in\mathbb{R}^m$, $v\in\mathbb{R}^n$ are nonzero vectors, then

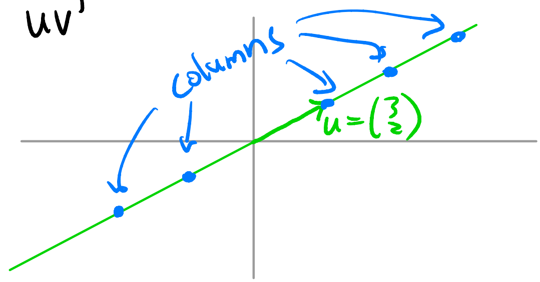
$$uv^T = u\underbrace{(v_1 \cdots v_n)}_{\text{weights}} = \begin{pmatrix} | & & | \\ v_1u & \cdots & v_nu \\ | & & | \end{pmatrix}$$

multiples of u

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

Let's plot the columns of uv^T (the data points).

$$\underbrace{\begin{pmatrix}3\\2\end{pmatrix}}_{u}\underbrace{\begin{pmatrix}-1 & 2 & 1 & 3 & -2\end{pmatrix}}_{v^T}$$



The columns are $-1\cdot u$, $2u$, $1\cdot u$, $3u$, $-2u$.

Upshot: A matrix A of rank 1 encodes data points (columns) that lie on a line ($\mathrm{Col}(A)$). The outer product decomposition $A=uv^T$ tells you

which line: $\mathrm{Span}\{u\}$

and which multiples of u: the entries of v^T.

$r=2$: In this case, $A = u_1 v_1^T + u_2 v_2^T$.

weights

$$u_1 v_1^T + u_2 v_2^T = u_1 \begin{pmatrix} v_{11} & \cdots & v_{1n} \end{pmatrix} + u_2 \begin{pmatrix} v_{21} & \cdots & v_{2n} \end{pmatrix}$$

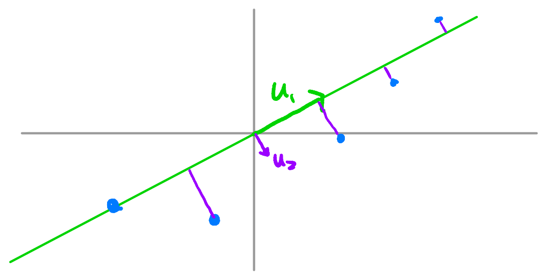
$$= \begin{pmatrix} | & & | \\ v_{11}u_1 + v_{21}u_2 & \cdots & v_{1n}u_1 + v_{2n}u_2 \\ | & & | \end{pmatrix}$$

linear combinations of u_1, u_2

This is an $m \times n$ matrix of rank 2: the columns are linear combinations of u_1, u_2, so

$$\mathrm{Col}(A) = \mathrm{Span}\{u_1, u_2\} \text{ is a plane.}$$

Let's plot the columns of A (the data points).



$$A = \overset{u_1}{\begin{pmatrix} 3 \\ 2 \end{pmatrix}} \overbrace{\begin{pmatrix} -1 & 2 & 1 & 3 & -2 \end{pmatrix}}^{v_1^T = \text{weights of } u_1} + \underset{u_2}{\begin{pmatrix} .2 \\ -.3 \end{pmatrix}} \underbrace{\begin{pmatrix} 3 & 1 & 2 & -1 & 0 \end{pmatrix}}_{v_2^T = \text{weights of } u_2}$$

orthogonal

Upshot: A matrix A of rank 2 encodes data points (columns) that lie on a plane ($\mathrm{Col}(A)$). The outer product decomposition $A = u_1v_1 + u_2v_2$ tells you

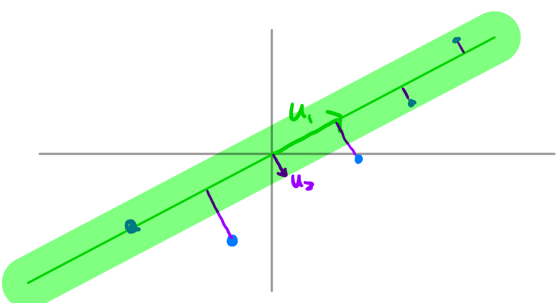
which plane: $\mathrm{Span}\{u_1, u_2\}$

and the weights of u_1, u_2: the entries of v_1^T, v_2^T.

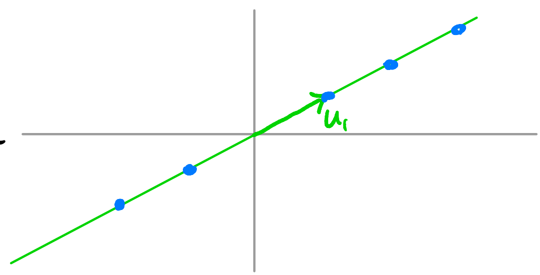
BUT: $\left\|\begin{pmatrix}3\\2\end{pmatrix}\right\| \gg \left\|\begin{pmatrix}.2\\-.3\end{pmatrix}\right\|$, so the $\begin{pmatrix}.2\\-.3\end{pmatrix}$-direction is less important!

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

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



$\approx$



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

In general, the SVD will find:

- the best-fit line
- the best-fit plane
- the best-fit 3-space

etc., for our data, all at once, and tell you how well they fit your data in the sense of orthogonal least squares. (L26, L27)

Why might you care?

- Data compression: if A is a 2×5 matrix and it almost has rank 1, then $A \approx u_1 v_1^T$.

$$A = \begin{pmatrix} \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \end{pmatrix} \text{ has 10 numbers, but}$$

$$u_1 v_1^T = \begin{pmatrix} \bullet \\ \bullet \end{pmatrix} \begin{pmatrix} \bullet & \bullet & \bullet & \bullet & \bullet \end{pmatrix} \text{ only has 7.}$$

- Data analysis: The SVD will reveal all approximate linear relations among your data points.
- Dimension Reduction: If our data points are in $\mathbb{R}^{1,000,000}$ but almost lie on a 100-dimensional subspace, then computers only need to use 100 numbers, not 1,000,000 (curse of dimensionality).
- Statistics: SVD finds important correlations.

etc...

Mechanics of the SVD

Recall the statement of the

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

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

where:

- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$
- $\{u_1, u_2, \dots, u_r\}$ is an orthonormal set in $\mathbb{R}^m$.
- $\{v_1, v_2, \dots, v_r\}$ is an orthonormal set in $\mathbb{R}^n$.

The quantities in the theorem all have names.

Def:

$$\left.\begin{array}{l} \bullet\ \sigma_1, \sigma_2, \dots, \sigma_r \text{ are the singular values} \\ \bullet\ u_1, u_2, \dots, u_r \text{ are the left singular vectors} \\ \bullet\ v_1, v_2, \dots, v_r \text{ are the right singular vectors} \end{array}\right\} \text{ of } A$$

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

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

$$
\begin{aligned}
Ax &= (\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T)x \\
&= \sigma_1 u_1 (v_1^T x) + \sigma_2 u_2 (v_2^T x) + \cdots + \sigma_r u_r (v_r^T x) \\
&= \sigma_1 u_1 (v_1 \cdot x) + \sigma_2 u_2 (v_2 \cdot x) + \cdots + \sigma_r u_r (v_r \cdot x)
\end{aligned}
$$

$$\Rightarrow \boxed{Ax = \sigma_1 (v_1 \cdot x) u_1 + \sigma_2 (v_2 \cdot x) u_2 + \cdots + \sigma_r (v_r \cdot x) u_r}$$

Formal Consequence ②: Taking $x = v_i$ above,

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

($\{v_1, v_2, \ldots, v_r\}$ is orthonormal)

Hence the singular vectors are related by:

$$\boxed{Av_i = \sigma_i u_i} \xrightarrow{\|u_i\| = 1} \boxed{\|Av_i\| = \sigma_i}$$

Formal Consequence ③:

$\{u_1, u_2, \ldots, u_r\}$ is an orthonormal basis for $\mathrm{Col}(A)$.

Indeed, ① shows that any $Ax \in \mathrm{Span}\{u_1, u_2, \ldots, u_r\}$, and ② shows $u_i = A\left(\frac{1}{\sigma_i} v_i\right) \in \mathrm{Col}(A)$.

Formal Consequence 4: Take transposes:

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

This is also an SVD — the only difference is we switched the u_i's and v_i's.

$$\boxed{\begin{array}{c}\text{The SVD of } A^T \text{ is} \\ A^T = \sigma_1 v_1 u_1^T + \sigma_2 v_2 u_2^T + \cdots + \sigma_r v_r u_r^T\end{array}}$$

In particular, A and A^T have the same:

- singular values $\sigma_1, \sigma_2, \ldots, \sigma_n$, and
- singular vectors (switch right and left).

Since $\mathrm{Col}(A^T) = \mathrm{Row}(A)$, 3 + 4 imply:

Formal Consequence 5:

$\{v_1, v_2, \ldots, v_r\}$ is an orthonormal basis for $\mathrm{Row}(A)$.

Formal Consequence 6:

Applying 2 and 4 gives

$$\boxed{A^T u_i = \sigma_i v_i} \quad \text{and} \quad \boxed{\|A^T u_i\| = \sigma_i}$$

Therefore,

$$A^TAv_i \overset{2}{=} A^T(\sigma_i u_i) = \sigma_i (A^Tu_i) \overset{\text{(above)}}{=} \sigma_i(\sigma_i v_i) = \sigma_i^2 v_i$$

$$AA^Tu_i \overset{\text{(above)}}{=} A(\sigma_i v_i) = \sigma_i (Av_i) \overset{2}{=} \sigma_i(\sigma_i u_i) = \sigma_i^2 u_i$$

$$\boxed{A^TAv_i = \sigma_i^2 v_i} \qquad \boxed{AA^Tu_i = \sigma_i^2 u_i}$$

This says:

$v_1, v_2, \dots, v_r$ are orthonormal eigenvectors of A^TA, with eigenvalues $\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2$

$u_1, u_2, \dots, u_r$ are orthonormal eigenvectors of AA^T, with eigenvalues $\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2$

This tells us how to prove the SVD exists / how to compute the SVD:

orthogonally diagonalize $S = A^TA$ or AA^T

Let's prove that the SVD exists.

Pay attention to steps 1-2: they illustrate the mechanics of the SVD.

Proof That the SVD Exists

Let $S = A^TA$. Recall that S is positive-semidefinite, so its eigenvalues are ≥ 0.

By the Spectral Theorem, $AM(\lambda) = GM(\lambda)$ for each eigenvalue λ, so I'll refer to both as the "multiplicity of λ".

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of S, in decreasing order. If an eigenvalue has multiplicity d, it appears d times in this list.

Step 1: I claim that 0 is an eigenvalue of multiplicity $n-r$ ($r = \mathrm{rank}(A)$).

(This just means 0 isn't an eigenvalue if $r=n$.)

Proof: The multiplicity of 0 is equal to

$$GM(0) = \dim \mathrm{Nul}(S - 0I_n) = \dim \mathrm{Nul}(S)$$
$$= \dim \mathrm{Nul}(A^TA).$$

But $\mathrm{Nul}(A^TA) \overset{(L10)}{=} \mathrm{Nul}(A)$ and $\dim \mathrm{Nul}(A) \overset{(L8)}{=} n-r$, so the multiplicity of 0 is $n-r$. //

Step 1 implies $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$ (zero is the smallest eigenvalue, so it comes last).

Therefore the nonzero eigenvalues of $S = A^T A$ are

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

Now we can define the singular values and the singular vectors.

$$\sigma_1 = \sqrt{\lambda_1} \qquad \sigma_2 = \sqrt{\lambda_2} \qquad \cdots \qquad \sigma_r = \sqrt{\lambda_r}$$

Let $\{v_1, v_2, \ldots, v_r\}$ be orthonormal eigenvectors of S with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$, respectively.

(This uses AM=GM again: if λ_1 has multiplicity 2 then $\lambda_1 = \lambda_2$ and there are two LI λ_1-eigenvectors.)

We know what the u_i's have to be:

$$u_1 = \frac{1}{\sigma_1} A v_1 \qquad u_2 = \frac{1}{\sigma_2} A v_2 \qquad \cdots \qquad u_r = \frac{1}{\sigma_r} A v_r$$

Step 2: I claim $\{u_1, u_2, \ldots, u_r\}$ is orthonormal.

Proof: $u_i \cdot u_j = \left(\frac{1}{\sigma_i} A v_i\right) \cdot \left(\frac{1}{\sigma_j} A v_j\right) = \left(\frac{1}{\sigma_i} A v_i\right)^T \left(\frac{1}{\sigma_j} A v_j\right)$

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

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

$$\overset{Sv_j = \sigma_j^2 v_j}{=} \frac{1}{\sigma_i \sigma_j} v_i^T (\sigma_j^2 v_j) = \frac{\sigma_j}{\sigma_i} v_i^T v_j$$

$$= \frac{\sigma_j}{\sigma_i} v_i \cdot v_j$$

Now we use the fact that $\{v_1, v_2, \ldots, v_r\}$ is orthonormal:

$i=j$: this $= \frac{\sigma_i}{\sigma_i} v_i \cdot v_i = 1$

$i \neq j$: this $= \frac{\sigma_j}{\sigma_i} v_i \cdot v_j = 0$ //

Now we know what all of the singular values and vectors are supposed to be, so the only thing left to do is:

Step 3: Verification that

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

Proof: Let $B = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$. We want to show that $A = B$. Recall (L11) that it's enough to show that $Ax = Bx$ for all vectors $x \in \mathbb{R}^n$.

Let $\{v_{r+1}, v_{r+2}, \ldots, v_n\}$ be an orthonormal basis for the (0-eigenspace of S) $= \mathrm{Nul}(S) = \mathrm{Nul}(A)$.

Then $\{v_1, v_2, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ is an orthonormal eigenbasis of S. (We just didn't do the 0 eigenvalue yet.)

(i) $Av_i = \sigma_i u_i = Bv_i \ (i \leq r)$:

$Av_i = \sigma_i u_i$ by definition of u_i.

$$Bv_i = (\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T) v_i$$

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

$= \sigma_i u_i$, as in Formal Consequence 2

(ii) $Av_i = 0 = Bv_i \ (i > r)$:

$Av_i = 0$ because $v_i \in \mathrm{Nul}(S) = \mathrm{Nul}(A)$.

$$Bv_i = (\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T) v_i$$

$$= \sigma_1 \underbrace{(v_1 \cdot v_i)}_{=0} u_1 + \sigma_2 \underbrace{(v_2 \cdot v_i)}_{=0} u_2 + \cdots + \sigma_r \underbrace{(v_r \cdot v_i)}_{=0} u_r$$

$= 0$ because v_i is $\perp v_1, v_2, \ldots, v_r$ $(r < i)$.

(iii) $Ax = Bx$ for any vector x:

Since $\{v_1, v_2, \ldots, v_n\}$ is a basis for $\mathbb{R}^n$, we can expand in the eigenbasis:

$$x = x_1 v_1 + x_2 v_2 + \cdots + x_n v_n$$

$$\Rightarrow Ax = A(x_1 v_1 + x_2 v_2 + \cdots + x_n v_n)$$

$$= x_1 Av_1 + x_2 Av_2 + \cdots + x_n Av_n$$

$$\overset{(i, ii)}{=} x_1 Bv_1 + x_2 Bv_2 + \cdots + x_n Bv_n$$

$$= B(x_1 v_1 + x_2 v_2 + \cdots + x_n v_n) = Bx \ //$$

Summary: Mechanics of the SVD

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

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

$$Ax = \sigma_1 (v_1 \cdot x) u_1 + \sigma_2 (v_2 \cdot x) u_2 + \cdots + \sigma_r (v_r \cdot x) u_r$$

Singular Values: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2$ are the nonzero eigenvalues of $A^T A$ and AA^T

Left singular Vectors: $\{u_1, u_2, \ldots, u_r\}$

$\rightarrow$ Orthonormal eigenvectors of AA^T:

$$\boxed{AA^T u_i = \sigma_i^2 u_i}$$

$\rightarrow$ Orthonormal basis for $\text{Col}(A)$

Right singular Vectors: $\{v_1, v_2, \ldots, v_r\}$

$\rightarrow$ Orthonormal eigenvectors of $A^T A$:

$$\boxed{A^T A v_i = \sigma_i^2 v_i}$$

$\rightarrow$ Orthonormal basis for $\text{Row}(A)$

The singular vectors are related by:

$\boxed{Av_i = \sigma_i u_i}$ $\quad$ $\boxed{A^T u_i = \sigma_i v_i}$ $\quad$ $\boxed{\|Av_i\| = \sigma_i = \|A^T u_i\|}$

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

NB: A^TA and AA^T have the

same nonzero eigenvalues $\sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2$.

(We showed in Formal Consequence 6 that these are eigenvalues of A^TA and AA^T, and we showed in the proof that the other eigenvalues $= 0$.)

Q: What about the 0 eigenvalue?

Hint: What if A is a tall matrix with FCR?

The proof also gives a procedure to compute the SVD (see below).

NB: This is not the algorithm used in practice! Efficiently computing the SVD is a hard problem. See the course website for some links to real-world algorithms.

NB: If A is wide $(m<n)$ then it's probably easier to compute the SVD of A^T:

A^TA is $n\times n$ but AA^T is $m\times m$,

so it's easier to find eigenvalues and eigenvectors of AA^T in this case.

Naive Schoolbook Procedure to Compute the SVD:

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

(1) Compute the nonzero eigenvalues of $S = A^T A$:

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

(The λ_i's appear multiple times if their multiplicities are ≥ 2.)

→ There are automatically r of them (counted with multiplicity), and they are positive.

(2) Find an orthonormal basis for the λ_i-eigenspace $(i = 1, 2, \ldots, r)$ ⇝ get an orthonormal set $\{v_1, v_2, \ldots, v_r\}$ with $Sv_i = \lambda_i v_i$

→ Since $AM(\lambda_i) = GM(\lambda_i)$, you automatically get r vectors.

(3) Set $\sigma_i = \sqrt{\lambda_i}$ and $u_i = \frac{1}{\sigma_i} A v_i$ $(i = 1, 2, \ldots, r)$.

Then $\{u_1, u_2, \ldots, u_r\}$ is orthonormal, and

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

is the SVD of A.

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

(1) $S = A^T A = \begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix}$

$$p(\lambda) = \det(S - \lambda I_2) = \lambda^2 - 50\lambda + 225$$
$$= (\lambda - 45)(\lambda - 5)$$

so $\lambda_1 = 45 \geq \lambda_2 = 5$

(2) Compute eigenspaces:

$\lambda = 45$ $\begin{pmatrix} -b \\ a-\lambda \end{pmatrix} = \begin{pmatrix} -20 \\ -20 \end{pmatrix} \rightsquigarrow v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$\lambda = 5$ $\begin{pmatrix} -b \\ a-\lambda \end{pmatrix} = \begin{pmatrix} -20 \\ 20 \end{pmatrix} \rightsquigarrow v_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

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

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

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

Check: $\|u_1\| = \frac{1}{\sqrt{10}} \sqrt{1^2 + 3^2} = 1$ $u_1 \cdot u_2 = 0$ ✓

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

SVD:

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

NB: You don't want to cancel the $\sqrt{5}$'s and $\sqrt{10}$'s here! You want to remember that $\sigma_1 = 3\sqrt{5}$ & $\sigma_2 = \sqrt{5}$.