The Singular Value Decomposition (SVD)

Matrix factorisations (continued)

The singular value decomposition (known mostly by its acronym, SVD) is yet another type of factorisation/decomposition..

Singular values

Definition 1 (Singular value)

Let $A \in \mathcal{M}_{mn}(\mathbb{R})$. The **singular values** of A are the real numbers

$$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n \geq 0$$

that are the square roots of the eigenvalues of $A^T A$

Singular values are real and nonnegative?

Recall that $\forall A \in \mathcal{M}_{mn}$, $A^T A$ is symmetric

Claim 1. Real symmetric matrices have real eigenvalues

Proof. $A \in \mathcal{M}_n(\mathbb{R})$ symmetric and $(\lambda, \mathbf{v})$ eigenpair of A, i.e, $A\mathbf{v} = \lambda\mathbf{v}$. Taking the complex conjugate, $\overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}}$

Since $A \in \mathcal{M}_n(\mathbb{R})$, $\overline{A} = A \qquad (z = \bar{z} \iff z \in \mathbb{R})$

So

$$A\bar{\mathbf{v}} = \overline{A}\bar{\mathbf{v}} = \overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$$

i.e., if $(\lambda, \mathbf{v})$ eigenpair, $(\bar{\lambda}, \bar{\mathbf{v}})$ also eigenpair

Still assuming $A \in \mathcal{M}_n(\mathbb{R})$ symmetric and $(\lambda, \mathbf{v})$ eigenpair of A and using what we just proved (that $(\bar{\lambda}, \bar{\mathbf{v}})$ also eigenpair), take transposes

$$
\begin{aligned}
A\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}} &\iff (A\bar{\mathbf{v}})^T = (\bar{\lambda}\bar{\mathbf{v}})^T \\
&\iff \bar{\mathbf{v}}^T A^T = \bar{\lambda}\bar{\mathbf{v}}^T \\
&\iff \bar{\mathbf{v}}^T A = \bar{\lambda}\bar{\mathbf{v}}^T \qquad [A \text{ symmetric}]
\end{aligned}
$$

Let us now compute $\lambda(\bar{\mathbf{v}} \bullet \mathbf{v})$. We have

$$
\begin{aligned}
\lambda(\bar{\mathbf{v}} \bullet \mathbf{v}) &= \lambda\bar{\mathbf{v}}^T\mathbf{v} = \bar{\mathbf{v}}^T(\lambda\mathbf{v}) \\
&= \bar{\mathbf{v}}^T(A\mathbf{v}) = (\bar{\mathbf{v}}^T A)\mathbf{v} \\
&= (\bar{\lambda}\bar{\mathbf{v}}^T)\mathbf{v} = \bar{\lambda}(\bar{\mathbf{v}} \bullet \mathbf{v}) \\
&\iff (\lambda - \bar{\lambda})(\bar{\mathbf{v}} \bullet \mathbf{v}) = 0
\end{aligned}
$$

We have shown

$$(\lambda - \bar{\lambda})(\bar{\mathbf{v}} \bullet \mathbf{v}) = 0$$

Let

$$\mathbf{v} = \begin{pmatrix} a_1 + ib_1 \\ \vdots \\ a_n + ib_n \end{pmatrix}$$

Then

$$\bar{\mathbf{v}} = \begin{pmatrix} a_1 - ib_1 \\ \vdots \\ a_n - ib_n \end{pmatrix}$$

So

$$\bar{\mathbf{v}} \bullet \mathbf{v} = (a_1^2 + b_1^2) + \cdots + (a_n^2 + b_n^2)$$

But $\mathbf{v}$ eigenvector is $\neq \mathbf{0}$, so $\bar{\mathbf{v}} \bullet \mathbf{v} \neq 0$, so

$$(\lambda - \bar{\lambda})(\bar{\mathbf{v}} \bullet \mathbf{v}) = 0 \iff \lambda - \bar{\lambda} = 0 \iff \lambda = \bar{\lambda} \iff \lambda \in \mathbb{R}$$

Claim 2. For $A \in \mathcal{M}_{mn}(\mathbb{R})$, the eigenvalues of $A^T A$ are real and nonnegative

Proof. We know that for $A \in \mathcal{M}_{mn}$, $A^T A$ symmetric and from previous claim, if $A \in \mathcal{M}_{mn}(\mathbb{R})$, then $A^T A$ is symmetric and real and with real eigenvalues

Let $(\lambda, \mathbf{v})$ be an eigenpair of $A^T A$, with $\mathbf{v}$ chosen so that $\|\mathbf{v}\| = 1$

Norms are functions $V \to \mathbb{R}_+$, so $\|A\mathbf{v}\|$ and $\|A\mathbf{v}\|^2$ are ≥ 0 and thus

$$\begin{aligned}
0 \leq \|A\mathbf{v}\|^2 &= (A\mathbf{v}) \bullet (A\mathbf{v}) = (A\mathbf{v})^T (A\mathbf{v}) \\
&= \mathbf{v}^T A^T A\mathbf{v} = \mathbf{v}^T (A^T A\mathbf{v}) = \mathbf{v}^T (\lambda\mathbf{v}) \\
&= \lambda(\mathbf{v}^T \mathbf{v}) = \lambda(\mathbf{v} \bullet \mathbf{v}) = \lambda\|\mathbf{v}\|^2 \\
&= \lambda
\end{aligned}$$

The singular value decomposition (SVD)

Theorem 2 (SVD)

$A \in \mathcal{M}_{mn}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$

Then there exists $U \in \mathcal{M}_m$ orthogonal, $V \in \mathcal{M}_n$ orthogonal and a block matrix $\Sigma \in \mathcal{M}_{mn}$ taking the form

$$\Sigma = \begin{pmatrix} D & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{pmatrix}$$

where

$$D = \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \in \mathcal{M}_r$$

such that

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

Definition 3

We call a factorisation as in Theorem 2 the **singular value decomposition** of A. The columns of U and V are, respectively, the **left** and **right singular vectors** of A

U and V^T are *rotation* or *reflection* matrices, Σ is a *scaling* matrix

Outer product form of the SVD

Theorem 4 (Outer product form of the SVD)

$A \in \mathcal{M}_{mn}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$, $\mathbf{u}_1, \ldots, \mathbf{u}_r$ and $\mathbf{v}_1, \ldots, \mathbf{v}_r$, respectively, left and right singular vectors of A corresponding to these singular values

Then

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

Computing the SVD (case of $\neq$ eigenvalues)

To compute the SVD, we use the following result

Theorem 5

Let $A \in \mathcal{M}_n$ symmetric, $(\lambda_1, \mathbf{u}_1)$ and $(\lambda_2, \mathbf{u}_2)$ be eigenpairs, $\lambda_1 \neq \lambda_2$. Then $\mathbf{u}_1 \bullet \mathbf{u}_2 = 0$

Proof of Theorem 5

$A \in \mathcal{M}_n$ symmetric, $(\lambda_1, \mathbf{u}_1)$ and $(\lambda_2, \mathbf{u}_2)$ eigenpairs with $\lambda_1 \neq \lambda_2$

$$\begin{aligned}
\lambda_1(\mathbf{v}_1 \bullet \mathbf{v}_2) &= (\lambda_1 \mathbf{v}_1) \bullet \mathbf{v}_2 \\
&= A\mathbf{v}_1 \bullet \mathbf{v}_2 \\
&= (A\mathbf{v}_1)^T \mathbf{v}_2 \\
&= \mathbf{v}_1^T A^T \mathbf{v}_2 \\
&= \mathbf{v}_1^T (A\mathbf{v}_2) \qquad [A \text{ symmetric so } A^T = A] \\
&= \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) \\
&= \lambda_2 (\mathbf{v}_1^T \mathbf{v}_2) \\
&= \lambda_2 (\mathbf{v}_1 \bullet \mathbf{v}_2)
\end{aligned}$$

So $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \bullet \mathbf{v}_2) = 0$. But $\lambda_1 \neq \lambda_2$, so $\mathbf{v}_1 \bullet \mathbf{v}_2 = 0$

Computing the SVD (case of $\neq$ eigenvalues)

If all eigenvalues of A^TA are distinct, we can use Theorem 5

1. Compute $A^TA \in \mathcal{M}_n$
2. Compute eigenvalues $\lambda_1, \ldots, \lambda_n$ of A^TA; order them as $\lambda_1 > \cdots > \lambda_n \geq 0$ ($>$ not $\geq$ since $\neq$)
3. Compute singular values $\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_n = \sqrt{\lambda_n}$
4. Diagonal matrix D in Σ is either in $\mathcal{M}_n$ (if $\sigma_n > 0$) or in $\mathcal{M}_{n-1}$ (if $\sigma_n = 0$)

5. Since eigenvalues are distinct, Theorem 5 $\implies$ eigenvectors are orthogonal set. Compute these eigenvectors in the same order as the eigenvalues
6. Normalise them and use them to make the matrix V, i.e., $V = [\mathbf{v}_1 \cdots \mathbf{v}_n]$
7. To find the $\mathbf{u}_i$, compute, for $i = 1, \ldots, r$,

$$\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$$

and ensure that $\|\mathbf{u}_i\| = 1$

Computing the SVD (case where some eigenvalues are $=$)

1. Compute $A^T A \in \mathcal{M}_n$
2. Compute eigenvalues $\lambda_1, \ldots, \lambda_n$ of $A^T A$; order them as $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$
3. Compute singular values $\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_n = \sqrt{\lambda_n}$, with $r \leq n$ the index of the last positive singular value
4. For eigenvalues that are distinct, proceed as before
5. For eigenvalues with multiplicity > 1, we need to ensure that the resulting eigenvectors are LI *and* orthogonal

Dealing with eigenvalues with multiplicity > 1

When an eigenvalue has (algebraic) multiplicity > 1, e.g., characteristic polynomial contains a factor like $(\lambda - 2)^2$, things can become a little bit more complicated

The proper way to deal with this involves the so-called Jordan Normal Form (another matrix decomposition)

In short: not all square matrices are diagonalisable, but all square matrices admit a JNF

Sometimes, we can find several LI eigenvectors associated to the same eigenvalue. Check this. If not, need to use the following

Definition 6 (Generalised eigenvectors)

$\mathbf{x} \neq \mathbf{0}$ **generalized eigenvector** of rank m of $A \in \mathcal{M}_n$ corresponding to eigenvalue λ if

$$(A - \lambda \mathbb{I})^m \mathbf{x} = \mathbf{0}$$

but

$$(A - \lambda \mathbb{I})^{m-1} \mathbf{x} \neq \mathbf{0}$$

Procedure for generalised eigenvectors

$A \in \mathcal{M}_n$ and assume λ eigenvalue with algebraic multiplicity k

Find $\mathbf{v}_1$, "classic" eigenvector, i.e., $\mathbf{v}_1 \neq \mathbf{0}$ s.t. $(A - \lambda\mathbb{I})\mathbf{v}_1 = \mathbf{0}$

Find generalised eigenvector $\mathbf{v}_2$ of rank 2 by solving for $\mathbf{v}_2 \neq \mathbf{0}$,

$$(A - \lambda\mathbb{I})\mathbf{v}_2 = \mathbf{v}_1$$

...

Find generalised eigenvector $\mathbf{v}_k$ of rank k by solving for $\mathbf{v}_k \neq \mathbf{0}$,

$$(A - \lambda\mathbb{I})\mathbf{v}_k = \mathbf{v}_{k-1}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ LI

Back to the normal procedure

With the LI eigenvectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ corresponding to λ

Apply Gram-Schmidt to get orthogonal set

For all eigenvalues with multiplicity > 1, check that you either have LI eigenvectors or do what we just did

When you are done, be back on your merry way to step 6 in the case where eigenvalues are all $\neq$

I am caricaturing a little here: there can be cases that do not work exactly like this, but this is general enough..

Applications of the SVD

Many applications of the SVD, both theoretical and practical..

1. Obtaining a unique solutions to least squares when $A^T A$ singular
2. Image compression

Least squares revisited

Theorem 7

Let $A \in \mathcal{M}_{mn}$, $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$. The least squares problem $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution $\tilde{\mathbf{x}}$ of minimal length (closest to the origin) given by

$$\tilde{\mathbf{x}} = A^+\mathbf{b}$$

where A^+ is the pseudoinverse of A

Definition 8 (Pseudoinverse)

$A = U\Sigma V^T$ an SVD for $A \in \mathcal{M}_{mn}$, where

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}, \text{ with } D = \mathsf{diag}(\sigma_1, \ldots, \sigma_r)$$

(D contains the nonzero singular values of A ordered as usual)

The **pseudoinverse** (or **Moore-Penrose inverse**) of A is $A^+ \in \mathcal{M}_{nm}$ given by

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

with

$$\Sigma^+ = \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{nm}$$

Compressing images

Consider an image (for simplicity, assume in shades of grey). This can be stored in a matrix $A \in \mathcal{M}_{mn}$

Take the SVD of A. Then the small singular values carry information about the regions with little variation and can perhaps be omitted, whereas the large singular values carry information about more "dynamic" regions of the image

Suppose A has r nonzero singular values. For $k \leq r$, let

$$A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

(so for $k = r$ we get the usual outer product form)