Singular value decomposition(SVD)

Matrix decompositions

1. Lower-upper(LU)
2. Cholesky
3. QR
4. eigen decomposition(Schur) $\mathbf{A} = \mathbf{Q}^{-1}\Lambda\mathbf{Q}$
5. *Singular value decomposition(SVD)*

Eigendecomposition

Let $\mathbf{A} \in \mathbb{R}^{n\times n} \leftarrow$ real square and symmetric.

$$\mathbf{A} = \mathbf{Q}^{-1}\Lambda\mathbf{Q}$$

1. $\Lambda = \mathrm{diag}(\lambda_1,\ \lambda_2,\ \ldots,\lambda_n)$: eigenvalues
2. $\mathbf{Q}$ is orthogonal: $\mathbf{Q}^T\mathbf{Q} = \widehat{\mathbf{1}}$
3. Columns of $\mathbf{Q}$ are eigenvectors

$$\mathbf{A}\vec{q}_k = \lambda_k\, \vec{q}_k, \qquad k = 1, \ldots, n$$

Singular value decomposition

Let $\mathbf{A} \in \mathbb{R}^{n\times m}$. Let $m \leqslant n$ (for clarity, not required)

$$\underset{n\times m}{\mathbf{A}} = \underset{n\times n}{\mathbf{U}} \cdot \underset{n\times m}{\mathbf{\Sigma}} \cdot \underset{m\times m}{\mathbf{V}^T}$$

1. $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_m)$: *singular values*
2. $\mathbf{U}^T\mathbf{U} = \widehat{\mathbf{1}} \leftarrow n \times n$; $\mathbf{V}^T\mathbf{V} = \widehat{\mathbf{1}} \leftarrow m \times m$
3. Factorization exists for any square/rectangular matrix $\mathbf{A}$ (proof: GvL§2.4)

SVD

Let $\mathbf{A} \in \mathbb{R}^{n\times m}$. Let $m \leqslant n$ (for clarity, not required)

$$\underset{n\times m}{\mathbf{A}} = \underset{n\times n}{\mathbf{U}} \cdot \underset{n\times m}{\mathbf{\Sigma}} \cdot \underset{m\times m}{\mathbf{V}^T}$$

1. $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1,\ \sigma_2,\ \cdots,\ \sigma_m)$

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_m$$

2. $\mathbf{U}^T\mathbf{U} = \widehat{\mathbf{1}} \leftarrow n \times n$; $\mathbf{V}^T\mathbf{V} = \widehat{\mathbf{1}} \leftarrow m \times m$
3. Columns of $\mathbf{U}\&\mathbf{V}$:

$$\mathbf{U} = [\vec{u}_1 \mid \vec{u}_2 \mid \cdots \mid \vec{u}_n], \qquad \vec{u}_k \in \mathbb{R}^n$$: left singular vectors

$$\mathbf{V} = [\vec{v}_1 \mid \vec{v}_2 \mid \cdots \mid \vec{v}_m], \qquad \vec{v}_k \in \mathbb{R}^m$$: right singular vectors

Singular values and singular vectors

Right-multiply the factorization, $\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T$, by $\mathbf{V}$:

$$\mathbf{AV} = \mathbf{U\Sigma}$$

Take the k-th column:

$$\mathbf{A}\vec{v}_k = \sigma_k \vec{u}_k \,, \qquad k = 1, \ldots, \underset{\min(m,n)}{m} \tag{1}$$

Transpose and right-multiply by $\mathbf{U}$:

$$\mathbf{A}^T \vec{u}_k = \sigma_k \vec{v}_k \,, \qquad k = 1, \ldots, m \tag{2}$$

Singular values and singular vectors

Combine (1) and (2):

$$\mathbf{A}\left(\frac{1}{\sigma_k}\mathbf{A}^T\vec{u}_k\right) = \sigma_k\vec{u}_k \quad \Rightarrow \quad \mathbf{A}\mathbf{A}^T\vec{u}_k = \sigma_k^2\vec{u}_k$$

$$\mathbf{A}^T\left(\frac{1}{\sigma_k}\mathbf{A}\vec{v}_k\right) = \sigma_k\vec{v}_k \quad \Rightarrow \quad \mathbf{A}^T\mathbf{A}\vec{v}_k = \sigma_k^2\vec{v}_k$$

- σ_k^2 are eigenvalues of $\mathbf{A}^T\mathbf{A}$
- $\vec{v}_k$ are eigenvectors of $\mathbf{A}^T\mathbf{A}$
- Note that $\dim \mathbf{A}\mathbf{A}^T = n$, so only m eigenvalues are σ_k^2.

Rank, range and nullspace

Consider (1): $\mathbf{A}\vec{v}_k = \sigma_k \vec{u}_k.$

Suppose that r values of σ_k are non-zero:

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > \underbrace{\sigma_{r+1} \geqslant \cdots \geqslant \sigma_m}_{=0}$$

- $\operatorname{rank} \mathbf{\Sigma} = r \quad \Rightarrow \quad \operatorname{rank} \mathbf{A} = r$
- $\operatorname{null}(\mathbf{A}) = \operatorname{span}(\vec{v}_{r+1}, \cdots, \vec{v}_m)$

  ядро $\mathbf{A}$: подпространство $\vec{x}$: $\mathbf{A}\vec{x} = 0$
- $\operatorname{ran}(\mathbf{A}) = \operatorname{span}(\vec{u}_1, \cdots, \vec{u}_r)$

  образ $\mathbf{A}$: линейная комбинация столбцов: $\forall \vec{x}\ \mathbf{A}\vec{x} \in \operatorname{range}$

Numerically, define a cutoff for small singular values.

Subspaces and projections

Let $\vec{x} \in \mathbb{R}^n$. $\widehat{P}$ is a *projector* onto a subspace $S \subset \mathbb{R}^n$ if

$$\widehat{P}\vec{x} \in S \quad \text{and} \quad (1 - \widehat{P})\vec{x} \in S^{\perp}$$

Let $\operatorname{rank} \mathbf{A} = r$.
Partition $\mathbf{U}$ and $\mathbf{V}$ matrices:

$$\mathbf{U} = [\, \underset{\overset{\leftrightarrow}{r}}{\mathbf{U}_r} \mid \underset{\overset{\leftrightarrow}{n-r}}{\widetilde{\mathbf{U}}_r} \,], \qquad \mathbf{V} = [\, \underset{\overset{\leftrightarrow}{r}}{\mathbf{V}_r} \mid \underset{\overset{\leftrightarrow}{m-r}}{\widetilde{\mathbf{V}}_r} \,]$$

SVD generates projectors onto characteristic subspaces of $\mathbf{A}$

Subspaces and projections

Projectors via SVD:

- $\mathbf{V}_r \mathbf{V}_r^T$: projection onto $\mathrm{null}(\mathbf{A})^\perp = \mathrm{ran}(\mathbf{A}^T)$
- $\widetilde{\mathbf{V}}_r \widetilde{\mathbf{V}}_r^T$: projection onto $\mathrm{null}(\mathbf{A})$
- $\mathbf{U}_r \mathbf{U}_r^T$: projection onto $\mathrm{ran}(\mathbf{A}^T)$
- $\widetilde{\mathbf{U}}_r \widetilde{\mathbf{U}}_r^T$: projection onto $\mathrm{ran}(\mathbf{A})^\perp = \mathrm{null}(\mathbf{A}^T)$

Thin SVD (a.k.a. economic SVD)

Note that $\mathbf{\Sigma}$ has a block structure:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 \\ \mathbf{0} \end{bmatrix}$$

where $\mathbf{\Sigma}_1 = \mathrm{diag}(\sigma_1, \cdots, \sigma_m)$.

Decompose $\mathbf{U} = [\underset{\overset{\leftrightarrow}{m}}{\mathbf{U}_1} \mid \underset{\overset{\leftrightarrow}{n-m}}{\mathbf{U}_2}] \updownarrow$n

Then,

$$\mathbf{A} = [\,\mathbf{U}_1 \,|\mathbf{U}_2\,] \cdot \begin{bmatrix} \mathbf{\Sigma}_1 \\ 0 \end{bmatrix} \cdot \mathbf{V}^T = \mathbf{U}_1 \cdot \mathbf{\Sigma}_1 \cdot \mathbf{V}^T$$

U_1 has orthogonal columns.

SVD expansion

Write $\mathbf{A}$ as a sum of rank-1 matrices:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \equiv \sum_{k=1}^{m} \sigma_k \vec{u}_k \vec{v}_k^T$$

This is exact. Can use it for approximations: approximate a given matrix with a matrix of smaller rank.

Eckart–Young theorem

Let $\mathrm{rank}(\mathbf{A}) = r$. Consider

$$\mathbf{A}_p = \sum_{k=1}^{p} \sigma_k \vec{u}_k \vec{v}_k^T$$

note: $p < r$ terms

Consider a set of rank-p matrices $\mathbf{B}$: $\mathrm{rank}(\mathbf{B}) = p$. Then,

$$\min_{\mathbf{B}:\mathrm{rank}\,\mathbf{B}=p} \|\mathbf{A} - \mathbf{B}\|_2 = \|\mathbf{A} - \mathbf{A}_p\|_2$$

$$= \sigma_{p+1}$$

proof: GvL §2.4

Recall that the matrix 2-norm is $\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})} \equiv \sigma_{\max}(\mathbf{A})$

Interpretations

SVD decomposes a linear mapping $x \longrightarrow \mathbf{A}x$:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{k=1}^{r} \sigma_k u_k v_k^T$$

$$\xrightarrow{x} \quad \mathbf{V}^T \quad \xrightarrow{\mathbf{V}^T x} \quad \mathbf{\Sigma} \quad \xrightarrow{\mathbf{\Sigma}\mathbf{V}^T x} \quad \mathbf{U} \quad \longrightarrow \mathbf{A}x$$

1. rotate the basis into *input directions* $\vec{v}_1, \cdots, \vec{v}_r$
2. scale the coefficients in the input basis by σ_k
3. rotate to the output basis $\vec{u}_1, \cdots, \vec{u}_r$

NB: input/output directions are different.

Interpretations

SVD decomposes a linear mapping $x \longrightarrow \mathbf{A}x$:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{k=1}^{r} \sigma_k u_k v_k^T$$

$$\xrightarrow{x} \quad \mathbf{V}^T \quad \xrightarrow{\mathbf{V}^T x} \quad \mathbf{\Sigma} \quad \xrightarrow{\mathbf{\Sigma}\mathbf{V}^T x} \quad \mathbf{U} \quad \longrightarrow \quad \mathbf{A}x$$

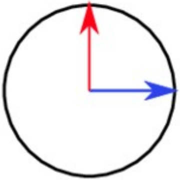
- v_1 is the most sensitive input direction
- u_1 is the highest gain output direction

Can $\approx$ neglect components which correspond to the smallest singular values.

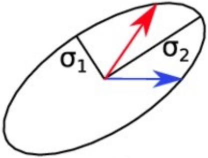
Properties of SVD

Geometrically:

circle $\|x\| \leq 1$



$\{Ax : \|x\| \leq 1\}$



$\vec{u_k}$: principal axes
σ_k: lengths of principal semiaxes

Geometric interpretation