MF TUTORIAL PART 1

Singular Value and Eigen-decompositions

Michiel Hochstenbach and Sibylle Hess

Norms

Frobenius norm $\|A\|^2 = \sum_{ij} A_{ij}^2$

Operator 2-norm $\|A\|_{\mathsf{op}} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$

Singular value decomposition (SVD): for any $A \in \mathbb{R}^{m \times n}$

$$A = U_r \Sigma_r V_r^T = \begin{bmatrix} \vdots & & \vdots \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \vdots & & \vdots \\ \mathbf{v}_1 & \cdots & \mathbf{v}_r \\ \vdots & & \vdots \end{bmatrix}^T$$

$r =$ rank

Best rank-k approximation $A \approx U_k \Sigma_k V_k^T$

SVD/PCA: Best Rank-k Approximation

TSVD: $A_k = U_k \Sigma_k V_k^T$

Solves:

$$A_k = \underset{\mathsf{rank}(B)=k}{\operatorname{argmin}} \|A - B\| \qquad \min = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_r^2}$$

$$A_k = \underset{\mathsf{rank}(B)=k}{\operatorname{argmin}} \|A - B\|_{\mathsf{op}} \qquad \min = \sigma_{k+1}$$

(not unique)

Useful for dimension reduction / feature extraction

SVD AND NORMS

Norm	Expression	
Frobenius	$\sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$	(= Schatten 2)
Operator 2	σ_1	(= Ky Fan 1)
Ky Fan	$\sigma_1 + \cdots + \sigma_\ell$	(any ℓ)
Nuclear	$\sigma_1 + \cdots + \sigma_r$	(or trace: tr$(A^T A)$)
Schatten p	$(\sigma_1^p + \cdots + \sigma_r^p)^{1/p}$	

Scientific computing / numerical linear algebra:

$$\mathbf{v}_1 = \underset{\|\mathbf{v}\|=1}{\operatorname{argmax}} \|A\mathbf{v}\|_F \qquad \sigma_1 = \|A\mathbf{v}_1\|$$

$$\mathbf{v}_2 = \underset{\|\mathbf{v}\|=1,\ \mathbf{v}\perp\mathbf{v}_1}{\operatorname{argmax}} \|A\mathbf{v}\|_F \qquad \sigma_2 = \|A\mathbf{v}_2\| \qquad \text{(etc)}$$

Statistics: maximal variability (variance) of centered data

SMALLEST SINGULAR VALUES

- Measure of distance to singular matrix (square), (data) matrix of lower rank (nonsquare)
- Plays role in sensitivity of solution to linear system $A\mathbf{x} = \mathbf{b}$
  Condition number $\kappa(A) = \sigma_1/\sigma_r$
- Dimension reduction: which features to EXclude first
  Not very common: $k \ll r$

SVD and Eigenvalue Decompositions

$A = U \Sigma V^T$

Three different but related eigenvalue decompositions:

- $A^T A = V \ \Sigma^T \Sigma \ V^T$
- $A A^T = U \ \Sigma \Sigma^T \ U^T$
- $\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} \Sigma$ (augmented matrix)

All of the involved operators are symmetric

Optimization

Computing partial SVD/PCA

Key message: computing partial SVD is one of easiest tasks in numerical linear algebra and may be VERY affordable !

- Small matrices, $m, n \leq 10000$: $\mathcal{O}(mn^2)$ operations
- Large-scale problems:
  APPLY A SEVERAL TIMES TO 1 (OR A FEW) VECTORS
  costs in terms of MVs (matrix-vector products)

  Number of MVs depends on distribution of singular values, but usually very modest

  One reason: $\text{gap} = \dfrac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 - \sigma_r^2} \approx 2 \cdot \dfrac{\sigma_1 - \sigma_2}{\sigma_1 - \sigma_r}$
- Huge-scale problems: randomized methods
  APPLY A ONCE TO A LOT OF RANDOM VECTORS

SVD/PCA for large-scale matrices

State-of-the-art: (Golub–Kahan–) Lanczos bidiagonalization

Krylov method:

- span$\{\mathbf{v},\ A\mathbf{v},\ (A^TA)\mathbf{v},\ A(A^TA)\mathbf{v},\ (A^TA)^2\mathbf{v},\dots\}$
- Short recurrences, but reorthogonalization necessary
- Therefore implicit restarts necessary
- Good paper/implementation: Baglama, Reichel 2005

Options:

- Largest σ's (by far most common!)
- Smallest σ's
- Smallest σ's $\neq 0$

SOFTWARE / CODES SVD

Python	`irlbpy`
Matlab	`svds`
R	`irlba`
Julia	(`irlba`?, in progress)

More extensive packages:

- PRIMME: eigenvalue solvers in C, Stathopoulos e.a.
  Has interfaces to Python and R
- SLEPc: parallel eigensolvers (built on PETSc), Roman e.a.
  Is interfaced by many packages, and has interfaces to ARPACK and BLOPEX
- BLOPEX: CG methods in C, Knyazev e.a.
- (ARPACK: F77 restarted Krylov, 1998), Sorensen e.a.
- (LAPACK: optimal implementations for small matrices)

SVD, NMF, Binary

$D \approx Y_k C_k X_k^T$

Decomp.	$\boldsymbol{X}_k$, $\boldsymbol{Y}_k$	$\boldsymbol{C}_k$
TSVD	orthogonal columns	diagonal PD
CUR	columns, rows **of** $\boldsymbol{D}$	full
NMF	positive elements	(diagonal PD)
Binary	$\{0,1\}$ or $\mathbb{1}$	full

See also next parts of tutorial by **Sibylle Hess**

Rank-1 SVD / NMF: (error ≈ 9.9)

$$\begin{bmatrix} 9 & 7 & 10 & 10 \\ 10 & 1 & 10 & 5 \\ 2 & 3 & 2 & 9 \\ 10 & 6 & 10 & 2 \end{bmatrix} \approx \begin{bmatrix} 0.63 \\ 0.51 \\ 0.26 \\ 0.52 \end{bmatrix} \begin{bmatrix} 28.2 \end{bmatrix} \begin{bmatrix} 0.59 \\ 0.31 \\ 0.61 \\ 0.43 \end{bmatrix}^T \approx \begin{bmatrix} 10.5 & 5.6 & 10.9 & 7.8 \\ 8.5 & 4.5 & 8.8 & 6.3 \\ 4.2 & 2.3 & 4.4 & 3.1 \\ 8.6 & 4.6 & 8.9 & 6.4 \end{bmatrix}$$

Rank-2 SVD: (error ≈ 4.2)

$$\begin{bmatrix} 0.63 & 0.33 \\ 0.51 & -0.29 \\ 0.26 & 0.75 \\ 0.52 & -0.49 \end{bmatrix} \cdot \begin{bmatrix} 28.2 & \\ & 9.0 \end{bmatrix} \cdot \begin{bmatrix} 0.59 & -0.37 \\ 0.31 & 0.15 \\ 0.61 & -0.33 \\ 0.43 & 0.86 \end{bmatrix}^T \approx \begin{bmatrix} 9.4 & 6.0 & 9.9 & 10.3 \\ 9.4 & 4.1 & 9.6 & 4.1 \\ 1.8 & 3.3 & 2.2 & 8.9 \\ 10.2 & 3.9 & 10.4 & 2.6 \end{bmatrix}$$

Rank-2 NMF: (error ≈ 9.3)

$$\begin{bmatrix} 0.76 & 0.15 \\ 0.20 & 0.96 \\ 0.40 & 0.15 \\ 0.48 & 0.17 \end{bmatrix} \cdot \begin{bmatrix} 10.80 & 8.09 \\ 10.03 & 0.00 \\ 11.32 & 8.20 \\ 12.20 & 1.52 \end{bmatrix}^T \approx \begin{bmatrix} 9.3 & 7.6 & 9.8 & 9.4 \\ 10.0 & 2.0 & 10.2 & 4.0 \\ 5.5 & 4.0 & 5.7 & 5.1 \\ 6.5 & 4.8 & 6.8 & 6.1 \end{bmatrix}$$

Rank-2 CUR: (error ≈ 18.9)

$$\begin{bmatrix} 9 & 10 \\ 10 & 5 \\ 2 & 9 \\ 10 & 2 \end{bmatrix} \begin{bmatrix} -0.03 & 0.12 \\ 0.12 & -0.12 \end{bmatrix} \begin{bmatrix} 2 & 10 \\ 3 & 6 \\ 2 & 10 \\ 9 & 2 \end{bmatrix}^T \approx \begin{bmatrix} 0.98 & 2.39 & 0.98 & 8.81 \\ 6.76 & 4.69 & 6.76 & 4.39 \\ -6.07 & -1.72 & -6.07 & 7.95 \\ 9.54 & 5.69 & 9.54 & 1.73 \end{bmatrix}$$

Some references part I

Lehoucq, Sorensen et al	1998	ARPACK: implicitly restarted Arnoldi
Stewart	2001	Krylov–Schur: implicitly restarted Krylov in an easy and elegant way
Baglama, Reichel	2005	Krylov–Schur for SVD
Halko, Martinsson et al	2011	Randomized methods
Sorensen, Embree	2016	CUR via DEIM

m.e.hochstenbach@tue.nl