Approximate Matrix Rank

Nicholas Richardson

15 April 2024

THE UNIVERSITY OF BRITISH COLUMBIA
Mathematics
Faculty of Science

institute of applied MATHEMATICS
University of British Columbia

Rank of a matrix

Rank of a matrix

- $\text{rank}(A) =$ # independent rows/colums of A

Rank of a matrix

- $\mathrm{rank}(A) =$ # independent rows/colums of A
- Compute # nonzero singular values σ

Rank of a matrix

- $\text{rank}(A) =$ # independent rows/colums of A
- Compute # nonzero singular values σ
- Represents the number of import components in A

Rank of a matrix

- $\mathrm{rank}(A) =$ # independent rows/colums of A
- Compute # nonzero singular values σ
- Represents the number of import components in A
- If the rank r is small, can greatly compress data:

$$\underbrace{\begin{bmatrix} \quad & & \\ & A & \\ & & \quad \end{bmatrix}}_{m \times n} = \underbrace{\begin{bmatrix} \\ B \\ \\ \end{bmatrix}}_{m \times r} \underbrace{\begin{bmatrix} \quad C \quad \end{bmatrix}}_{r \times n}$$

What if there is noise?

$$\underset{m\times n}{\left[\quad \hat{A} \quad\right]} = \underset{m\times r}{\left[\, B \,\right]} \left[\; \underset{r\times n}{C} \;\right] + \underset{m\times n}{\left[\quad N \quad\right]}$$

What if there is noise?

- Gaussian noise matrix N is full rank

$$\underbrace{\left[\quad \hat{A} \quad\right]}_{m\times n} = \underbrace{\left[\, B \,\right]}_{m\times r}\underbrace{\left[\quad C \quad\right]}_{r\times n} + \underbrace{\left[\quad N \quad\right]}_{m\times n}$$

What if there is noise?

- Gaussian noise matrix N is full rank
- The noisy data $\hat{A} = A + N$ is also full rank

$$\underset{m \times n}{\begin{bmatrix} & & \\ & \hat{A} & \\ & & \end{bmatrix}} = \underset{m \times r}{\begin{bmatrix} \\ B \\ \\ \end{bmatrix}} \underset{r \times n}{\begin{bmatrix} & C & \end{bmatrix}} + \underset{m \times n}{\begin{bmatrix} & & \\ & N & \\ & & \end{bmatrix}}$$

What if there is noise?

- Gaussian noise matrix N is full rank
- The noisy data $\hat{A} = A + N$ is also full rank
- Don’t know the number of features in your data 🙁

$$\underset{m \times n}{\begin{bmatrix} & & \\ & \hat{A} & \\ & & \end{bmatrix}} = \underset{m \times r}{\begin{bmatrix} \\ B \\ \\ \end{bmatrix}} \underset{r \times n}{\begin{bmatrix} & C & \end{bmatrix}} + \underset{m \times n}{\begin{bmatrix} & & \\ & N & \\ & & \end{bmatrix}}$$

You need... ε-rank![1]

$$\mathrm{rank}_\varepsilon(A) = \min_{\|E\| \le \varepsilon} \mathrm{rank}(A + E)$$

1. Golub, Klema, and Stewart (1977)

You need... ε-rank![1]

$$\operatorname{rank}_\varepsilon(A) = \min_{\|E\| \le \varepsilon} \operatorname{rank}(A + E)$$

- Smallest rank of an ε-approximation of A

1. Golub, Klema, and Stewart (1977)

You need... ε-rank![1]

$$\operatorname{rank}_\varepsilon(A) = \min_{\|E\| \leq \varepsilon} \operatorname{rank}(A + E)$$

- Smallest rank of an ε-approximation of A
- Counts the # singular values *bigger* than ε

1. Golub, Klema, and Stewart (1977)

You need... ε-rank![1]

$$\operatorname{rank}_\varepsilon(A) = \min_{\|E\| \le \varepsilon} \operatorname{rank}(A + E)$$

- Smallest rank of an ε-approximation of A
- Counts the # singular values *bigger* than ε
- A.K.A. numerical or approximate rank

1. Golub, Klema, and Stewart (1977)

Useful feature included!

1. Vershynin (2018)

Useful feature included!

- If $\|N\| \leq \varepsilon < \sigma_r/2$,

$$\operatorname{rank}_\varepsilon(A+N) = \operatorname{rank}(A).$$

1. Vershynin (2018)

Useful feature included!

- If $\|N\| \leq \varepsilon < \sigma_r/2$,

$$\mathrm{rank}_\varepsilon(A+N) = \mathrm{rank}(A).$$

- σ_r is the smallest nonzero singular value of A

1. Vershynin (2018)

Useful feature included!

- If $\|N\| \leq \varepsilon < \sigma_r/2$,

$$\mathrm{rank}_\varepsilon(A+N) = \mathrm{rank}(A).$$

- σ_r is the smallest nonzero singular value of A
- ε needs to be between the noise level $\|N\|$ and the smallest important data (σ_r)

1. Vershynin (2018)

Useful feature included!

- If $\|N\| \leq \varepsilon < \sigma_r/2$,

$$\mathbf{rank}_\varepsilon(A + N) = \mathbf{rank}(A).$$

- σ_r is the smallest nonzero singular value of A
- ε needs to be between the noise level $\|N\|$ and the smallest important data (σ_r)
- Estimate standard Gaussian[1] $\|N\| \approx \sqrt{m} + \sqrt{n}$

1. Vershynin (2018)

Illustration of ε-rank

Illustration of ε-rank

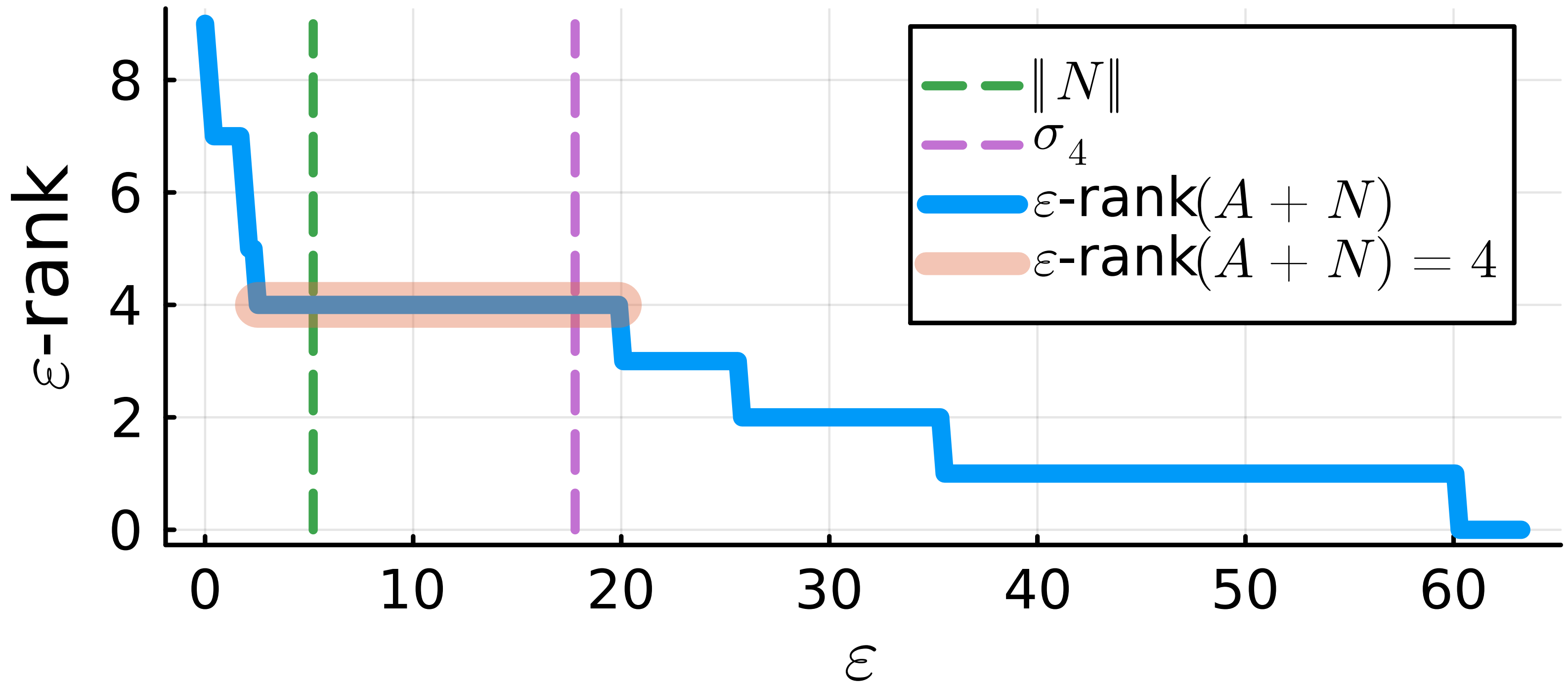
- Generate random 9×9 rank-4 matrix A

Illustration of ε-rank

- Generate random 9×9 rank-4 matrix A
- Add standard Gaussian noise N

Illustration of ε-rank

- Generate random 9×9 rank-4 matrix A
- Add standard Gaussian noise N



Future Work & References

Golub, Gene H., Virginia Klema, and Gilbert W. Stewart. 1977. "Rosetak Document 4: Rank Degeneracies and Least Square Problems." SSRN Scholarly Paper. Rochester, NY.

Vershynin, Roman. 2018. *High-Dimensional Probability*. Cambridge Series in Statistical and Probabilistic Mathematics. University of California, Irvine: Cambridge University Press.

Future Work & References

- How to pick the right ε?

Golub, Gene H., Virginia Klema, and Gilbert W. Stewart. 1977. "Rosetak Document 4: Rank Degeneracies and Least Square Problems." SSRN Scholarly Paper. Rochester, NY.

Vershynin, Roman. 2018. *High-Dimensional Probability*. Cambridge Series in Statistical and Probabilistic Mathematics. University of California, Irvine: Cambridge University Press.

Future Work & References

- How to pick the right ε?
- What if A and $A + N$ are nonnegative?

Golub, Gene H., Virginia Klema, and Gilbert W. Stewart. 1977. "Rosetak Document 4: Rank Degeneracies and Least Square Problems." SSRN Scholarly Paper. Rochester, NY.

Vershynin, Roman. 2018. *High-Dimensional Probability*. Cambridge Series in Statistical and Probabilistic Mathematics. University of California, Irvine: Cambridge University Press.