Introduction to Machine Learning

Singular Value Decomposition

Varun Chandola

Computer Science & Engineering
State University of New York at Buffalo
Buffalo, NY, USA
chandola@buffalo.edu

University at Buffalo
Department of Computer Science and Engineering
School of Engeering and Applied Sciences

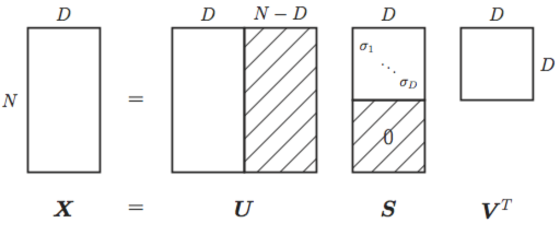
Outline

Decomposing a Matrix

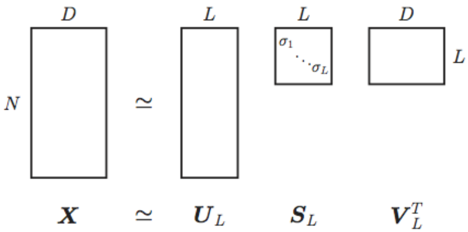
- For any matrix $\mathbf{X}$ $(N \times D)$

$$\underbrace{\mathbf{X}}_{N \times D} = \underbrace{\mathbf{U}}_{N \times N} \underbrace{\mathbf{S}}_{N \times D} \underbrace{\mathbf{V}^\top}_{D \times D}$$

Singular Value Decomposition - Illustration



(a)



(b)

Economy Sized or Thin SVD

- Assume that $N > D$

$$\underbrace{\mathbf{X}}_{N\times D} = \underbrace{\tilde{\mathbf{U}}}_{N\times D}\underbrace{\tilde{\mathbf{S}}}_{D\times D}\underbrace{\tilde{\mathbf{V}}^\top}_{D\times D}$$

Connection between Eigenvectors and Singular Vectors

- Let $\mathbf{X} = \mathbf{USV}^\top$

$$\begin{aligned} \mathbf{X}^\top\mathbf{X} &= \mathbf{VS}^\top\mathbf{U}^\top\mathbf{USV}^\top \\ &= \mathbf{V}(\mathbf{S}^\top\mathbf{S})\mathbf{V}^\top \\ &= \mathbf{VDV}^\top \end{aligned}$$

- where $\mathbf{D} = \mathbf{S}^2$ is a diagonal matrix containing squares of singular values.
- Hence,

$$(\mathbf{X}^\top\mathbf{X})\mathbf{V} = \mathbf{VD}$$

- Which means that the columns of $\mathbf{V}$ are the eigenvectors of $\mathbf{X}^\top\mathbf{X}$ and $\mathbf{D}$ contains the eigenvalues.
- Similarly one can show that the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{XX}^\top$ and $\mathbf{D}$ contains the eigenvalues.

PCA Using SVD

- Assuming that $\mathbf{X}$ is centered (zero mean) the principal components are equal to the right singular vectors of $\mathbf{X}$.

Low Rank Approximations Using SVD

- Choose only first L singular values

$$\underbrace{\mathbf{X}}_{N\times D} \approx \underbrace{\tilde{\mathbf{U}}}_{N\times L} \underbrace{\tilde{\mathbf{S}}}_{L\times L} \underbrace{\tilde{\mathbf{V}}^\top}_{L\times D}$$

- Only need $NL + LD + L$ values to represent $N \times D$ matrix
- Also known as *rank L approximation* of the matrix $\mathbf{X}$ (Why?)

The Matrix Approximation Lemma

- Among all possible rank L approximations of a matrix $\mathbf{X}$, SVD gives the best approximation
  - In the sense of minimizing the *Frobenius norm*

$$\|\mathbf{X} - \mathbf{X}_L\|$$

- Also known as the `Eckart Young Mirsky` theorem

Equivalence Between PCA and SVD

- For data $\mathbf{X}$ (assuming it to be centered)
- Principal components are the eigenvectors of $\mathbf{X}^\top\mathbf{X}$
- Or, principal components are the columns of $\mathbf{V}$

$$\mathbf{W} = \mathbf{V}$$

- Or

$$\hat{\mathbf{W}} = \hat{\mathbf{V}}$$

- $\hat{\mathbf{W}}$ are the first L principal components and $\hat{\mathbf{V}}$ are the first L right singular vectors.

Optimal Reconstruction for PCA

- For PCA, data in latent space:

$$\begin{aligned}\hat{\mathbf{Z}} &= \mathbf{X}\hat{\mathbf{W}} \\ &= \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^\top\mathbf{V} \\ &= \hat{\mathbf{U}}\hat{\mathbf{S}}\end{aligned}$$

- Optimal reconstruction for PCA:

$$\begin{aligned}\hat{\mathbf{X}} &= \hat{\mathbf{Z}}\hat{\mathbf{W}}^\top \\ &= \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^\top\end{aligned}$$

- **Optimal reconstruction is same as truncated SVD approximation!!**

Singular Value Decomposition - Recap

- What is the (column) rank of a matrix?
- For $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ (SVD):
  - What is the rank of $\hat{\mathbf{X}}^{(1)} = \mathbf{U}_{:1}\sigma_1\mathbf{V}_{:1}^\top$?
- How much storage is needed for a rank 1 matrix?
  - $O(N)$

Importance of the Matrix Approximation Lemma

- There are many ways to "approximate" a matrix with a lower rank approximation
- Low rank approximation allows us to *store* the matrix using much less than $N \times D$ bits ($O(N \times L)$ bits only)
- SVD gives the *best possible* approximation

$$\|\mathbf{X} - \hat{\mathbf{X}}\|_2^2$$

Why SVD?

- A faster way to do PCA (truncated SVD, sparse SVD)
- Other applications as well:
  - Image compression
  - Recommender Systems
    - There are better methods
  - Topic modeling (Latent Semantic Indexing)

References