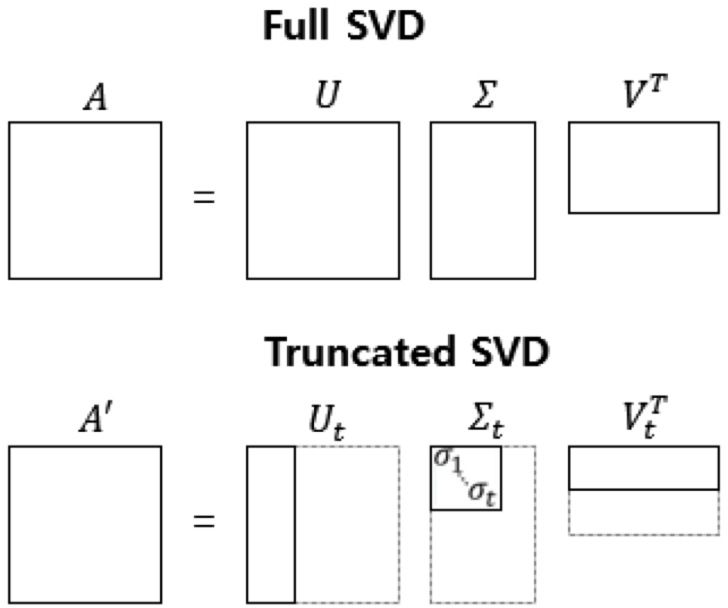
Latent Semantic Analysis (LSA)

Index

- PPMI
- LSA via Matrix Factorization
- Word2Vec via Matrix Factorization

LSA via Matrix Factorization

- Use ‘truncated SVD(Singular Value Decomposition)’ as a factorization method.



LSA via Matrix Factorization

- Can lower the dimension of original Matrix.
- This could effectively contains latent/hidden meaning of original matrix.
- It also reduce noise, sparsity of the input data(original matrix)

Word2Vec via Matrix Factorization

- Word2Vec = SPMI(Shifted PMI)

$$SPMI(A,B) = PMI(A,B) - \log k$$

$$\underset{\text{Word2Vec}}{U_i \cdot V_j} = \underset{\text{SPMI}}{PMI(i,j)} - \log k$$