

# 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.

**Full SVD**

$$\begin{array}{c} A \\ \square \end{array} = \begin{array}{c} U \\ \square \end{array} \begin{array}{c} \Sigma \\ \square \end{array} \begin{array}{c} V^T \\ \square \end{array}$$

**Truncated SVD**

$$\begin{array}{c} A' \\ \square \end{array} = \begin{array}{c} U_t \\ \begin{array}{|c|} \hline \square \\ \hline \end{array} \end{array} \begin{array}{c} \Sigma_t \\ \begin{array}{|c|} \hline \sigma_1 \backslash \sigma_t \\ \hline \end{array} \end{array} \begin{array}{c} V_t^T \\ \begin{array}{|c|} \hline \square \\ \hline \end{array} \end{array}$$

# 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)

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

$$U_i \cdot V_j = PMI(i, j) - \log k$$

Word2Vec                      SPMI