**Team 13**

Members: 2

Shu Xia, [sxia38@wisc.edu](mailto:sxia38@wisc.edu)

Ziqian Deng, zdeng67@wisc.edu

**Project**: Topic Modeling with SVD and NMF

**Introduction**

In nowadays business, we often come across data that doesn’t have outputs or labels. We strive to find the hidden structure and latent groupings within the unstructured content such as queries, social media, videos, customer reviews. This is when Natural Language Processing, one of the most essential unsupervised machine learning techniques comes into play.

Topic modeling is a fun way to start our study of NLP. In this activity, we will introduce the basics of topic modeling, combine with two popular matrix decomposition techniques --- Singular Value Decomposition and Non-negative Matrix Factorization, and compare these two techniques with each other.

**Goal**

After going through this activity, students will be able to:

* Have a knowledge of the wide range of applications in NLP
* Understand what is topic modeling, and learn how to represent topic modeling problems using matrix method.
* Refresh their knowledge in SVD and apply SVD to topic modeling problems
* Learn the Non-negative Matrix Factorization (NMF) and apply NMF to topic modeling problems
* Compare these two matrix factorization techniques and explain the different characteristics.

Recourses:

Topic Modeling

1. Intro video to latent semantic analysis (topic modeling):

<https://www.youtube.com/watch?v=hB51kkus-Rc>

NMF:

1. Video lecture on NMF for text classification: <https://www.youtube.com/watch?v=ZTxXGZwe2gw&list=WL&index=12&t=0s>
2. Nature article on NMF and its application to face recognition and text classification:

* Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755), 788-791.

1. NMF algorithm review article:

* Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P., & Plemmons, R. J. (2007). Algorithms and applications for approximate nonnegative matrix factorization. Computational statistics & data analysis, 52(1), 155-173.

1. NMF tools: sklearn documentation on NMF

<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF.html>