Algorithms and Applications of Data Mining

Yijun Lin

yijunlin@usc.edu

02/20

Clustering

- What is the curse of dimensionality?

- Distance measurement methods
  - Cosine, Jaccard, Euclidean
  - E.g., comparing two paragraphs, which to use?

- Hierarchical Clustering
  - How to represent a cluster of many points in Euclidean space?
  - How to determine "nearness" of clusters?

- K-Means Clustering
  - https://spark.apache.org/docs/latest/mllib-clustering.html

Dimensionality Reduction

- What is the rank of a matrix?

- Why dimensionality reduction? Applications?

- Singular Value Decomposition
    - $A = U \times \sum \times V^T$

    https://www.youtube.com/watch?v=EokL7E6o1AE

    -