

CS780 Midterm Topics:

1. Introduction
 - a. Distance metric - requirements
2. Time Series
 - a. Representations
 - i. Lower-bounding
 - ii. DFT, Wavelet, PAA, SAX, iSAX, APCA, PLA, SVD
 - b. Distance Measures
 - i. Shape-based
 1. Euclidean Distance
 2. Dynamic Time Warping
 3. LCSS
 - ii. Structural similarity
 1. Compression-based
 2. histogram-based (bag-of-patterns)
 3. model-based
 - c. Clustering
 - i. Whole clustering vs. subsequence clustering
 - d. Motif discovery
 - i. Random projection
 - ii. VizTree
 - iii. Grammar-based motif discovery
3. Text Mining
 - a. Indexing
 - i. Inverted index
 - ii. Vector space model
 - iii. LSI/SVD
 - b. Text classification
 - i. Naïve bayes
 - ii. KNN
 - iii. Rocchio
4. Web Mining
 - a. PageRank algorithm
 - b. HITS algorithm