

Some Review Questions:

Questions:

- 1) What is the main issue if we aim to learn topics from a corpus of documents by simply performing clustering?
- 2) Compare pLSI with LDA.
- 3) Cosine similarity is well known measure for computing similarity between any vector in a space
 - a) Show whether this measure is symmetric
 - b) Determine the output range
 - c) Show whether its output is invariant to shift in input domain.
 - d) Show whether its output is invariant to input negation (assuming inputs are binary).
- 4) Compare Cosine similarity with Pearson correlation, and under which circumstance both measures are equally likely (assuming input range is \mathbb{R})