Parameter Dimensionality

1 Data Simulation

We simulated a sparse dataset using LDA generative process. Parameter settings are as follows: the corpus is a collection of 5 documents and the vocabulary has 10 unique words. The number of words per document follows Poisson distribution with parameter 30. Each document is a mixture of 3 topics and the Dirichlet prior α for topics proportions is 0.001. Each topic is a mixture over the vocabulary and the Dirichlet prior η for the topic-specific word distribution is 0.01. we fitted the LDA model over 500 iterations with two random seeds and plot their complete log-likelihoods (Figure 1).
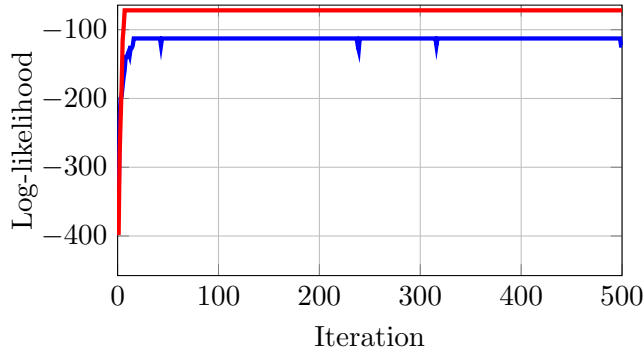
Figure 1: Left: Plot of log-likelihoods of the LDA model over 500 iterations with two random seeds, α equals 0.001 and η equals 0.01.