Choosing the Number of Topics in LDA Models – A Monte Carlo Comparison of Selection Criteria

Victor Bystrov, Viktoriia Naboka, Anna Staszewska-Bystrova, Peter Winker

Abstract

The performance of the recently developed singular Bayesian information criterion (sBIC) in selecting the number of topics in LDA models is evaluated and compared to the performance of alternative model selection criteria proposed for topic models. The sBIC is a generalization of the standard BIC that can be implemented to singular statistical models. The comparison is based on Monte Carlo simulations and carried out for several alternative settings, varying with respect to the number of topics, the number of documents and the size of documents in the corpora. Practical recommendations for LDA model selection are developed.