

# NIPS data analysis

Miaoyan Wang, Jan 14, 2021

The NIPS dataset [?] records the word counts from 2,483 papers published from 1987 to 2003. We focus on the top 100 authors, 200 most frequent words (excluding the first 100), and normalize the count by log transformation with pseudocount 1. We organize the resulting data into an order-3 continuous-valued tensor  $\mathcal{Y}$ , where the entry  $\mathcal{Y}(i, j, k)$  represents the log count of word  $i$  by author  $j$  in year  $k$ .

## References