Text Mining Datasets

NIPS paper dataset from UCI ML repo ($n < p$)

5811 documents with word counts for 11463 different words from NIPS text mining dataset at UCI-repo, corresponding to NIPS dataset in "Poisson Random Fields for Dynamic Feature Models"
This dataset consisting of 5811 documents from each of the dataset from NIPS.
We obtain 5811 documents from each category there are 5811 rows, each row represents one document;
We counts the word frequencies for each document, there are 11463 columns, each column represents one distinct word.

- `NIPS_counts.csv` stores the word count matrix of size 5811*11463.
- `NIPS_tfidf.csv` stores the TF-IDF matrix of size 5811*11463.
  There are no labels for this datasets, but each file is approximately 1.7GB with dense matrix forms.