Document Classification

Text Classification: definition

- **Input:**
  - a document d
  - a fixed set of classes $C = \{c_1, c_2, \ldots, c_j\}$

- **Output: a predicted class $c \in C$**

Requirements

- **Any Data**
- **Any algorithms including Naive Bayes**
  - Implement Naive Bayes
    - Do not use library of Naive Bayes
- **Compare three (or more) algorithms**

Format

Student ID:	Name:	Score:

NLP Project #1 Document Classification

Methods

1) Naïve bayes
    a. Algorithm explanation
    b. Pseudo code for algorithms
    c. Code implementation (core part)
2) …
3) …

Please use at least 3 methods for classfier (Naïve bayes with binarization / Naïve bayes with n-gram / Naïve bayes with TF-IDF / Wordembedding / RNN …)

Experimental Setting

1) Dataset (statistical information, train/dev/test, etc.)
2) Evaluation measure
3) Baselines

Experimental Result

1) Overall performance
    a. Analysis
2) Qualitative results (2 or 3 samples for the performance comparison)

Conclusion

Q&A

- Thank you