

Machine Learning for NLP

Text Classification and Naïve Bayes

Text Classification:

- Assigns categories to text documents.

Naïve Bayes Classifier:

- Based on Bayes Rule with the assumption of feature independence.
- Multinomial and Boolean variants.

Advantages:

- Fast, low storage, robust to irrelevant features.

Evaluation Metrics

- Precision: % of selected items that are correct.
- Recall: % of correct items selected.
- F1 Score: Harmonic mean of precision and recall.

Part-of-Speech (POS) Tagging

Assigns grammatical categories to words.

Methods:

- Rule-based
- HMM
- MEMM, CRF

Challenges:

- Word ambiguity
- Handling unknown words

Machine Learning for NLP

Performance:

- About 97% accuracy with supervised models.