

UE23CS352A: Machine Learning Lab

Week 12: Naive Bayes Classifier

| | | |
|-------------------|---------------|-----------|
| Navya G N | PES2UG23CS372 | Section F |
| Date : 01-11-2025 | | |

Objective

The objective of this lab is to build and evaluate text classification models using different machine learning techniques. The primary tasks performed include:

1. Part A: Implementing a Multinomial Naive Bayes (MNB) classifier completely from scratch.
2. Part B : Tuning and evaluating a Sklearn Multinomial Naive Bayes classifier using hyperparameter optimization.
3. Part C : Implementing a Bayes Optimal Classifier (BOC) model using word embeddings and clustering to classify documents

The dataset used is PubMed 20k, which contains biomedical research abstracts categorized into classes.

Methodology

Multinomial NB from Scratch

- Performed text preprocessing: Tokenization, Lowercasing, Stop-word removal, Vocabulary creation
- Calculated class priors using training label proportions.
- Computed conditional word probabilities with Laplace smoothing.
- Predicted classes by computing log posterior probabilities.

Sklearn Multinomial NB

- Used TF-IDF vectorization for text representation.
- Used GridSearchCV to tune the alpha smoothing parameter.
- Evaluated using F1-score on test data.

Bayes Optimal Classifier

- Compute posterior probabilities for each class from the trained Naive Bayes model.
- Consider multiple hypotheses (parameter settings).
- Use Bayesian Model Averaging (BMA) to determine the optimal class label.
- Output the class with highest combined posterior probability.

Results and Analysis

- Part A: Screenshot of final test Accuracy, F1 Score and Confusion Matrix.

| ==== Test Set Evaluation (Custom Count-Based Naive Bayes) ==== | | | | |
|--|-----------|--------|----------|---------|
| Accuracy: 0.7483 | | | | |
| | precision | recall | f1-score | support |
| BACKGROUND | 0.54 | 0.57 | 0.55 | 3621 |
| CONCLUSIONS | 0.61 | 0.70 | 0.66 | 4571 |
| METHODS | 0.83 | 0.85 | 0.84 | 9897 |
| OBJECTIVE | 0.53 | 0.51 | 0.52 | 2333 |
| RESULTS | 0.88 | 0.78 | 0.83 | 9713 |
| accuracy | | | 0.75 | 30135 |
| macro avg | 0.68 | 0.69 | 0.68 | 30135 |
| weighted avg | 0.76 | 0.75 | 0.75 | 30135 |

Macro-averaged F1 score: 0.6809



- Part B: Screenshot of best hyperparameters found and their resulting F1 score.

```

Training initial Naive Bayes pipeline...
Training complete.

==== Test Set Evaluation (Initial Sklearn Model) ====
Accuracy: 0.7266
      precision    recall   f1-score   support
BACKGROUND       0.64     0.43     0.51      3621
CONCLUSIONS      0.62     0.61     0.62      4571
METHODS          0.72     0.90     0.80     9897
OBJECTIVE         0.73     0.10     0.18     2333
RESULTS           0.80     0.87     0.83     9713

accuracy           -         -     0.73      30135
macro avg          0.70     0.58     0.59      30135
weighted avg        0.72     0.73     0.70      30135

Macro-averaged F1 score: 0.5877

Starting Hyperparameter Tuning on Development Set...
Grid search complete.

Best Parameters: {'nb_alpha': 0.1, 'tfidf_min_df': 3, 'tfidf_ngram_range': (1, 3)}
Best Cross-Validation F1 Score: 0.7044

```

- Part C:

1. Screenshot of SRN and sample size.

```

Please enter your full SRN (e.g., PES1UG22CS345): PES2UG23CS372
Using dynamic sample size: 10372
Actual sampled training set size used: 10372

```

2. Screenshot of BOC final Accuracy, F1 Score and Confusion Matrix.

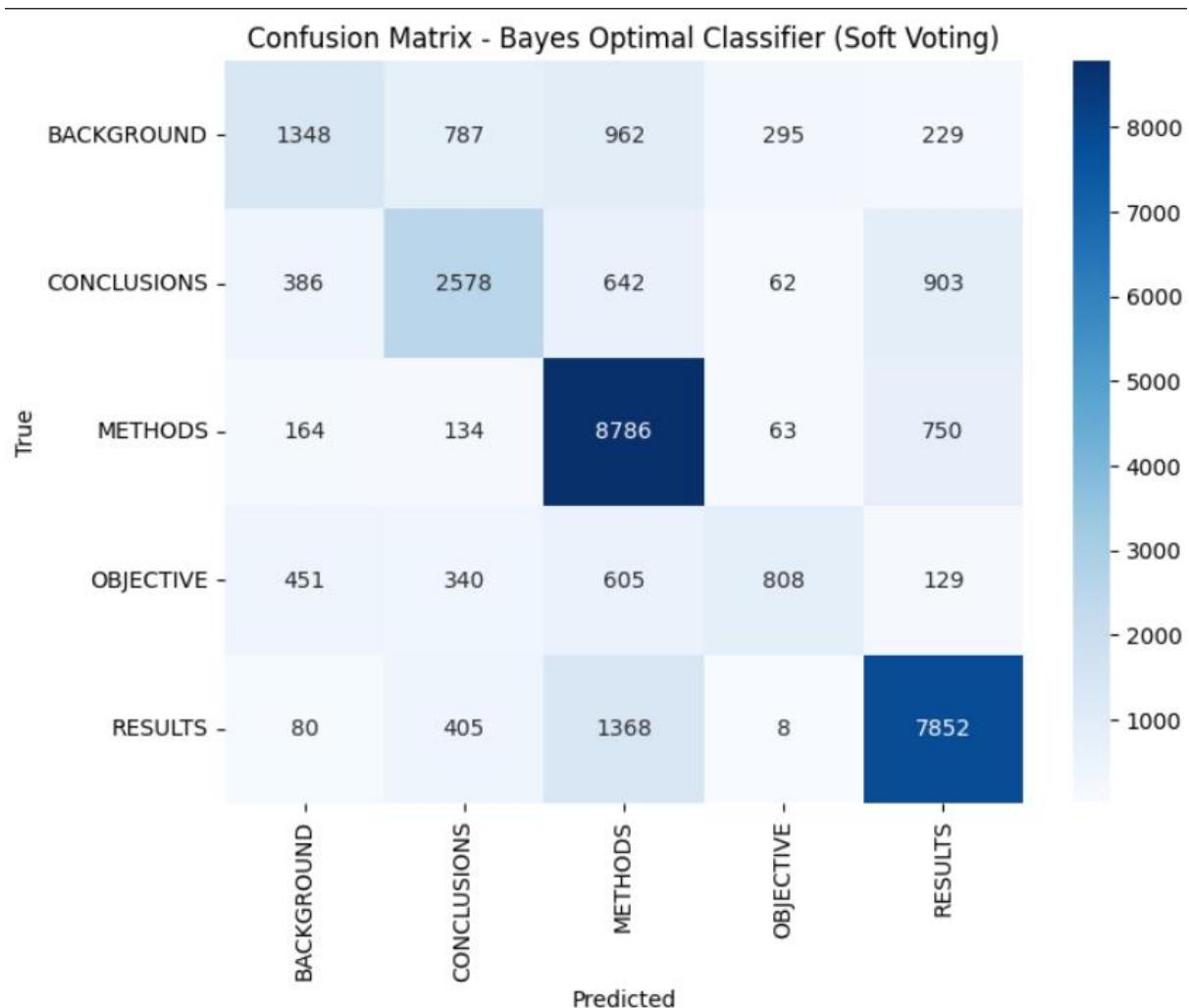
```

==== Final Evaluation: Bayes Optimal Classifier (Soft Voting) ====
Accuracy: 0.7092
Macro F1 : 0.6150

Classification Report:
      precision    recall   f1-score   support
BACKGROUND       0.55     0.37     0.45      3621
CONCLUSIONS      0.61     0.56     0.58      4571
METHODS          0.71     0.89     0.79     9897
OBJECTIVE         0.65     0.35     0.45     2333
RESULTS           0.80     0.81     0.80     9713

accuracy           -         -     0.71      30135
macro avg          0.66     0.60     0.61      30135
weighted avg        0.70     0.71     0.70      30135

```



- Discussion: Compare the performance of your scratch model (Part A) vs. the tuned Sklearn model (Part B) vs. the BOC approximation (Part C).

Part A — Multinomial Naive Bayes from Scratch

Test Accuracy: **0.7483**

Macro F1: **0.6809**

Part B - Sklearn MultinomialNB and Hyperparameter Tuning

Test Accuracy: **0.7266**

Macro F1: **0.5877**

Best CV f1_macro: **0.7044**

Best Parameters: {'nb_alpha': 0.1, 'tfidf_min_df': 3, 'tfidf_ngram_range': (1, 3)}

Part C - Bayes Optimal Classifier

BOC Accuracy: **0.7092**

BOC Macro F1: **0.6150**

Overall, the scratch implementation (Part A) performed the best on the test data, followed by BOC (Part C) and then the tuned sklearn model (Part B). This highlights that while hyperparameter tuning and word embeddings are valuable, the effectiveness of Naive Bayes still depends heavily on the nature of feature representation and data distribution.