

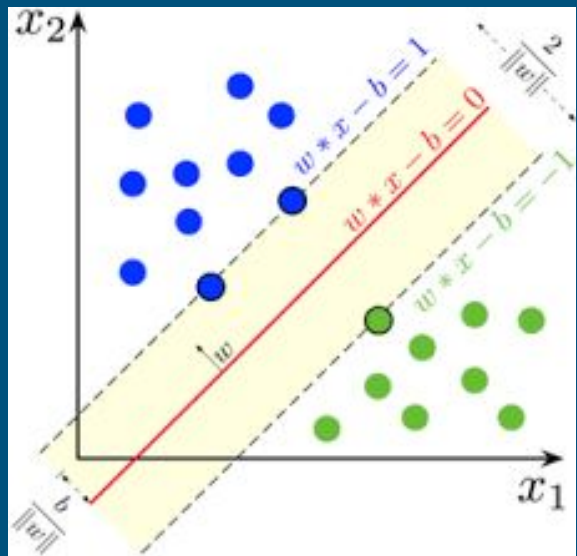
Sentiment Analysis on IMDB movie review using SVM

Faculty-
Prof. Dr. P. Aparna

By-
Mayank Rajpurohit (202SP015)
Sameer Koleshwar (202SP010)

Introduction

- What is Sentiment Analysis ?
- IMDB's movie review dataset
 - 50,000 movie reviews.
 - 25,000 Positive reviews
 - 25,000 Negative reviews.
- Support Vector Classifier



Pre-processing of dataset

- Tokenization
- Removal of stopwords
- Stemming
- Transformation:
 - Bag of Words
 - Term Frequency - Inverse Document Frequency (TF - IDF)
 - $TF = \text{Frequency of word in sentence} / \text{Total no. of words in sentence}$
 - $IDF = \log(\text{Total no. of sentences} / \text{No. of sentences contains that word})$
 - n-gram

Support Vector Classifier (SVC)

- SVMs classify an input vector $x \in R_d$ using:

$$y_i = \operatorname{argmax} (w^T x)$$

- Optimization problem:

$$\text{Min } 0.5 * ||w||^2$$

$$\text{Subject to } (x_i \cdot w + b) \cdot y_i \geq 1 \text{ for all } i$$

- Crammer - Singer formulation for optimization problem

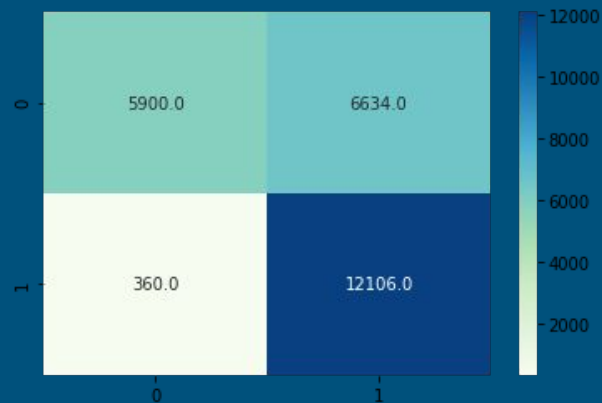
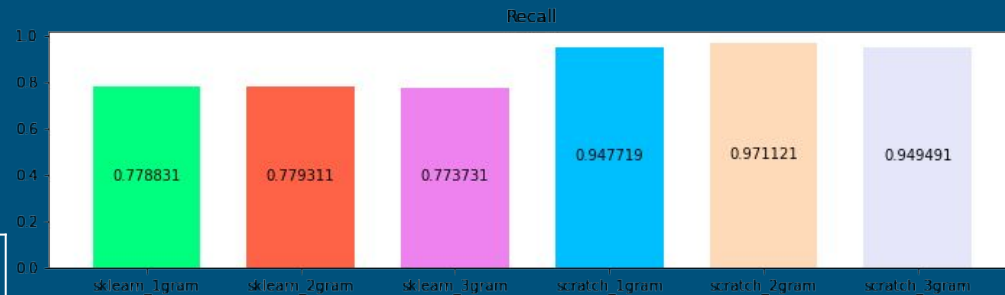
$$\text{minimize } (0.5 * \sum ||w_m||^2) + C (\sum [1 + \max (w_m^T \cdot x_i - w_{y_i}^T \cdot x_i)])$$

Algorithm for solving Crammer-Singer equation

1. INPUT: X as train data, Y as train labels
2. $X = \text{hstack}([X, \text{ones}((\text{len}(X), 1))])$
3. $N, d = X.\text{shape}$
4. $k = \text{no. of classes}$
5. Initialize W matrix of size (k, d) , penalty constant $C = 1.0$, $i=0$
6. while $i < \text{iterations}$ do
 - a. $\text{grad} = \text{gradient}(W, X, Y, C)$
 - b. $W = W - \alpha * \text{grad}$ (Gradient Descent Approach)
 - c. $i = i + 1$
7. end

Results

Models	Precision	Recall	F1-Score	Accuracy
1-Gram Sklearn	0.772	0.778	0.775	0.773
2-Gram Sklearn	0.788	0.779	0.784	0.786
3-Gram Sklearn	0.792	0.773	0.783	0.785
1-Gram Scratch	0.681	0.948	0.792	0.753
2-Gram Scratch	0.646	0.971	0.776	0.720
3-Gram Scratch	0.683	0.945	0.795	0.756



References

- [1] Aditi Sharan Sheeba Naz and Nidhi Malik. “ “sentiment classification on twitter data using support vector machine””. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI) (2018).
- [2] Casi Setianingsih Rimba Nuzulul Chory Muhammad Nasrun. “Sentiment analysis on user satisfaction level of monile data services using support vector machine (SVM) algorithm”. In: IEEE International Conference on Internet of Things and Intelligence System (2018).
- [3] Andia Enggar Mayasari and Anggit Dwi Hartanto. “User Satisfaction Levels Sentiment Analysis Toward Goods Delivery Service On Twitter Using Support Vector Machine Algorithm (SVM)”. In: International Conference on Information Technology, Information system and Electrical Engineering (2019).
- [4] Yoram Singer Koby Crammer. “On the Algorithmic Implementation of Multiclass Kernelbased Vector Machines”. In: Journal of Machine Learning Research (2001).

Thank you