

# Text multi-classification based on SoftMax Regression

## Project Proposal

Tianyi Xiong

### 1. Introduction and Motivation

Text classification is a popular topic in NLP. There are lots of methods for text classification such as Naïve Bayes, Logistic Regression, TF-IDF, Count Features, etc. In this project I will use SoftMax Regression to classify text, I will apply SoftMax on Bag of word model and N-Gram model to compare which method has higher efficiency and accuracy.

### 2. Previous work and Database.

This task is like PA2(logistic regression) to some extent. It will contain loss update and weight update like PA2, and it also have something new like different update strategy which could be selected by ourselves (Like stochastic and batch). For the database documents, I download the movie review document from the Rotten Tomatoes database, basically, it is about sentiment Analysis on movie review, each sentence has a label which represents the sentiment of this sentence (use number 0 to 5, and 0 represents negative and 5 represents positive). Here is the database link: <https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews/overview/description>

### 3. Deliverables

The basic target of this project is using a well-trained weight matrix from train class to predict the label (sentiment) of the sentence in the test class and calculate the accuracy.

Also, the loss diagrams will be provided.

Presentation slides

Final report.

Maybe there will be some other interesting functions (I have not decided yet)