MATH 456 — Mathematical Modeling

Project #3: Sentiment Analysis of Covid-19

The coronavirus disease 2019 (COVID-19) global pandemic has been a catastrophic event with major impact on the world's economy which created rise in unemployment, psychological issues, and depression. During rise of COVID-19 cases, and stricter lock downs, people have been expressing different sentiments in social media such as Twitter. Social media has played a significant role during COVID-19 which has driven researchers for analysis with NLP and machine learning methods. The research on the public's sentiments is essential for keeping mental health and informed about Covid-19.

Data. The training data contains 5000 labeled tweets while the released validation data have 2500 pieces of unlabeled tweets. The training data have 3 columns, containing Tweet ID, Tweet text, and labels. Note that the orders are shown as Optimistic (0), Thankful (1), Empathetic (2), Pessimistic (3), Anxious (4), Sad (5), Annoyed (6), Denial (7), Surprise (8), Official report (9), Joking (10). For example, if the labels are 3 and 6, it means that this piece of the tweet is labeled as Pessimistic and Annoyed.

Goal. Build a mathematical model for sentiment analysis via tweets. You may want to test your prediction of sentiments by using the validation dataset. However, notice that the validation dataset does not contain a score. You are recommended to use few lines (e.g. 50 lines) of the training set as the test data. You may first assign scores subjectively to tweets in the validation dataset and then compare the subjective scores with the predicted scores based on your model.

Numerous models and algorithms for sentiments analysis of tweets are available in the literature. You are allowed to use and adapt any model in the literature. However, you are responsible for **the originality of your codes and the following:**

- formulating the model(s) in a precise mathematical language;

- developing and implementing algorithms for computing your model(s);

- documenting and discussing your results in a short report (less than 10 pages), and presenting the results as a group.