Gebze Technical University
Department of Computer Engineering
CSE 654 / 484
Fall 2020

Homework 03
Due date: Jan 31st 2021

In this homework we will use word2vec representation of words with Naïve Bayes to make sentiment analysis of the Turkish movie reviews at
https://www.kaggle.com/ozcan15/turkish-sentiment-analysis-data-beyazperdecom
 Here are the steps of the homework

1. Download the word2vect source code https://github.com/tmikolov/word2vec . Compile and run it on the Beyazperde set (train and test combined). Run a few distance and analogy demos to see if the vectors are fine.
2. Use the dataset that you used for HW2 to train your word vectors. Include the Beyazperde set in your training.
3. Train a Naïve Bayes sentiment classifier on the training set. However, you will not use any counts of word to calculate the probability of P(w | c) where c is the class and w is the word. You will use word vectors to make this estimation. An important part of your homework will be figuring out this probability calculation. You should define at least two ways of making this calculation.

Prepare your report and submit it to the moodle page. You may use any programming language for the implementation. The word2vec source code in C language but you will use it to produce only the word vectors.

Notes
1. You will demo your homework result online
2. Your report should be very clear about the formula for P calculation