# E0 334 - Deep Learning for NLP

## Assignment 1

### (due by 16th Aug, 11:59 PM)

Note: Use the following link for submitting your results of Assignment 1.

`https://forms.office.com/r/8Ejdmtq1h5`

**Problem:**

The aim of this assignment is to study the use of different pre-trained word embeddings (word2vec/GloVe/fastText) for text representation and use them for text classification. You can download pre-trained word embeddings of your choice [1, 2, 3].

A simple way to represent a sentence is to tokenize, vectorize and average the word vector representations of all the words in the sentence. You will study the use of this simple approach of text representation for solving a text classification problem.

1. Use t-SNE [4] to plot the text representations obtained using the above-mentioned method. Indicate different classes with different colours. Study the use of different dimensional word embeddings in obtaining text representations.
2. Design a binary classifier using Deep Averaging Network (DAN)[5] for the dataset provided. Use the above-mentioned form to submit your results.

Note: You can use libraries like Gensim (`https://radimrehurek.com/gensim/`) or spaCy (`https://spacy.io/`) for various text processing as well as other tasks.

**References**


1. Word2vec (`https://code.google.com/archive/p/word2vec/`)
2. GloVe (`https://nlp.stanford.edu/projects/glove/`)
3. fastText (`https://fasttext.cc/`)
4. t-SNE Software (`https://lvdmaaten.github.io/tsne/`)
5. Iyyer *et al*, Deep Unordered Composition Rivals Syntactic Methods for Text Classification. `https://people.cs.umass.edu/~miyyer/pubs/2015_acl_dan.pdf`. Also see the code available at `https://github.com/miyyer/dan`