Fall 2018, CS 416: Algorithms for Machine Learning

Friday, October 27, 2018. Total Marks: 100 Points

Note: Each student is required to submit his own code and is expected to explain it to class **RA**.

Assignment Description:

1. scrap a data set of roman urdu which has atleast 10,000 unique words. [20]

2. Describe the process of scraping, provide urls used for scrapping and develop plain corpus [10]

3. Run skip gram word embedding algorithm with: [35]

   (a) k=5,10,15,20

   (b)embedding dimension=100,200,300.

4. visualize the embedding using tsne algorithm for [35]

   (a) 100 synonym pair words

   (b) 20 analogies

   (c) 100 same words but misspelled.