**CS671A: Introduction to NLP**
**Assignment #3: Dependency parsing**

Due on: 18-4-2018, 23.59                                                                 3-4-2018
MM: 250

1. In this assignment you have to build a transition based dependency parser (use the shift, left-arc, right-arc transition operations discussed in class).

   The key step is constructing the classifier (or oracle) that predicts the next transition operation $t$ given the current configuration.

   Invent a feature vector for a configuration. The choice of features you want to use is up to you. Build a neural network (or other) classifier that predicts the transition operation given a feature vector representation of the configuration $c$.

   To train the classifier you will have to use the dependency tagged treebank at:

   http://universaldependencies.org/ (use the EWT treebank for English)

   First, you have to use the treebank to generate pairs (configuration, operation) based on the trees in the treebank. The configuration should then be converted to a feature vector using your invented features so that the pairs become:

   (configuration-feature-vector, operation)

   The set of pairs above become your training set. Use this training set to train your neural n/w classifier (if you think you can do better using any other multi-class classifier then you are free to use that).

   You can set aside a small test set (say 5%) from the EWT data set (chosen randomly) to test your dependency parser.                                                                 [250]