Shiraz University

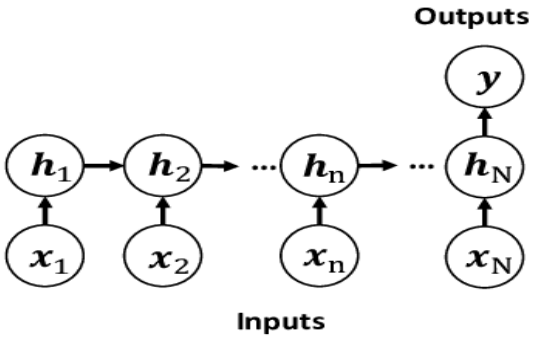
Neural Networks and Deep Learning (Ordibehesht 1401)

Assignment #2: Text Classification with Recurrent Neural Networks (RNN)

Due date: 6th Khordad 1401

RNNs are a type of Neural Networks where the output from the previous step is fed as input to the current step. RNN's are mainly used for Sentiment Analysis, Sequence Labeling, Speech Tagging, etc. There are four commonly used types of RNNs and one of them is Many-to-One.

Many-to-One is used when a single output is required from multiple input units or a sequence of them. It takes a sequence of inputs to predict a fixed output. Sentiment Analysis is a common example of this type of RNN. In order to do this, we need to consider a sentence as a word sequence (Many) and then predict its class label (One). That is the process of Many-to-One model. In the following figure, you can see a standard Many-to-One Architecture:



Part A: Document Preprocessing

1. Read ‘IMDB_Dataset.csv’ file and for each document:
   a. Remove all non-letter characters.

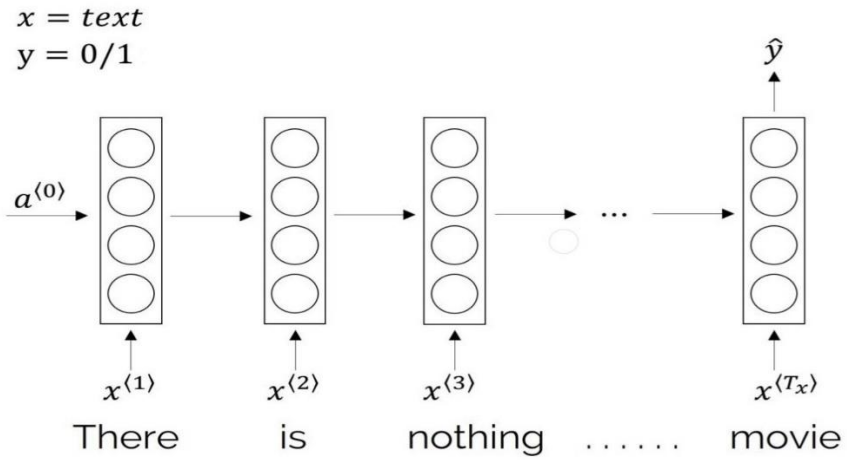
b. Remove the short words (length ≤ 2).
c. Remove all stop words (e.g., ‘a’, ‘and’, ‘what’, …), given in file ‘stopwords.txt’.

2. Tokenize the data and convert the text to word sequences.
3. Add padding to ensure that all the sequences have the same length (taking the max length).
4. Convert words of each sequence into numerical vector (use one-hot encoding vector).

Part B: Many-to-One model

1. Build Many-to-One RNN model with simple RNN (Elman Network).
2. Split documents into training and test data (80% for train).
3. Train your model and report the Accuracy on test data.

- Example of Many-to-One Architecture for Sentiment Analysis:



Dataset:

IMDB dataset has 5K movie reviews for natural language processing or text analytics and labeled by sentiment (positive and negative).

Notes:

- **Pay extra attention to the due date. It will not extend.**
- **Be advised that submissions after the deadline would not grade.**
- **Prepare your full report in PDF format and include the figures and results.**
- **Feel free to use any predefined functions.**
- **Email your files as a folder in this format (HW#_student#_name_family.zip).**
- **Email: pshiravani@gmail.com**