**CS 5316 – Natural Language Processing**

**Quiz 4 Solution**

**(Time limit: 12 minutes)**

1. (3 points) Given a corpus of positive and negative documents, explain a measure that can identify words with positive connotation.

We can use the odd-ratio or log odds-ratio to identify words conveying positive sentiment as follows:

P(w | c=+)/P(w | c=-) > t (t > 1)

1. (2 points) Attach appropriate POS tags to the words in the following sentence:

Islamabad/Noun beats/Verb Lahore/Noun by/Preposition 3//Number wickets/Noun in/Preposition a/Article tight/Adjective match/Noun

1. (3 points) Given a sequence model where and are input and target sequences of length , write down the expression for the best target sequence for the input sequence .
2. (2 points) Describe a method for aggregating sentiments associated with a specific aspect of a product (e.g., battery of a mobile phone).

* Prepare a list of words defining the aspect of interest
* Use a sentiment lexicon containing positive and negative words (e.g., SentiStrength)
* Define a window around the aspect word ( e.g., the sentence containing the word)
* Count the number of positive and negative words occurring within the window in the corpus
* Present the result as a graph or percentage