Feature extraction from text and images
TOTAL POINTS 4

1.Question 1
TF-IDF is applied to a matrix where each column represents a word, each row represents a document, and each value shows the number of times a particular word occurred in a particular document. Choose the correct statements.

(**X**) TF normalizes sum of the row values to 1

(**X**) IDF scales features inversely proportionally to a number of word occurrences over documents

() TF normalizes sum of the column values to 1

() IDF scales features proportional to the frequency of word's occurrences

1 point

2.Question 2
What of these methods can be used to preprocess texts?

() Plumbing

(**X**) Stopwords removal

() Levenshteining

(**X**) Lemmatization

() Plumping

(**X**) Lowercase transformation

(**X**) Stemming

1 point

3.Question 3
What is the main purpose of Lemmatization and Stemming?

() To remove words which are not useful.

() To induce common word amplification standards to the most useful for machine learning algorithms form.

() To reduce significance of common words.

(**X**) To reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.

1 point

4.Question 4
To learn Word2vec embeddings we need ...

() Labels for the documents in the corpora

() GloVe embeddings

() Labels for each word in the documents in the corpora

(**X**) Text corpora

1 point