✓ Congratulations! You passed!

Next Item



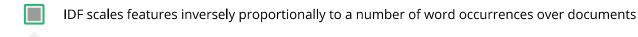
1/1 point

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.

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

Un-selected is correct



Correct

Correct! Purpose of IDF is to decrease importance of most frequent words.

TF normalizes sum of the column values to 1

Un-selected is correct

TF normalizes sum of the row values to 1

Correct

Correct!



1/1 point

2.

What of these methods can be used to preprocess texts? Feature extraction from text and images Practice Quiz, 4 Questivation Correct Correct! Lemmatization is used to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. Lowercase transformation Correct Correct! Lowercase is used to unite words written in UPPERCASE and lowercase. Levenshteining **Un-selected is correct** Stemming Correct Correct! Stemming is used to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. **Plumbing Un-selected is correct Plumping Un-selected is correct**

Correct

Correct! After feature generation stopwords result in useless features that interfere with useful ones.

Stopwords removal

Un-selected is correct

What is the main purpose of Lemmatization and Stemming?	
	To remove words which are not useful.
	To reduce significance of common words.
	To induce common word amplification standards to the most useful for machine learning algorithms form.
	To reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.
Correct You are right, this is the exact purpose of Lemmatization and Stemming!	
~	1/1 point
4. To learn Word2vec embeddings we need	
	Text corpora
	ect rect. To learn more about word embeddings, I encourage you to check out the course on Deep rning in our specialization.
	Labels for the documents in the corpora
Un-selected is correct	
	Labels for each word in the documents in the corpora
Un-selected is correct	
	GloVe embeddings

Feature extraction from text and images Practice Quiz, 4 questions



