coursera

Question 1

Select true statements about n-grams.

Correct answers:

- N-grams can help utilize local context around each word. Correct, because ngrams encode sequences of words.
- N-grams features are typically sparse. Correct. Ngrams deal with counts of words occurrences, and not every word can be found in a document. For example, if we count occurrences of words from an english dictionary in our everyday speech, a lot of words won't be there, and that is sparsity.

Incorrect answers:

- N-grams always help increase significance of important words. No, ngrams deals with words occurrences and not their importance.
- Levenshteining should always be applied before computing n-grams. Although, there is Levenshtein distance, there is no such thing as Levenshteining.

Question 2

Select true statements.

Correct answers:

- Bag of words usually produces longer vectors than Word2vec. Correct! Number of features in Bag of words approach is usually equal to number of unique words, while number of features in w2v is restricted to a constant, like 300 or so.
- Semantically similar words usually have similar word2vec embeddings. Correct. This is one of the main benefits of w2v in competitions.

Incorrect answers:

- Meaning of each value in BOW matrix is unknown. Incorrect. Meaning of a value in BOW matrix is the number of a word's occurrences in a document.
- You do not need bag of words features in a competition if you have word2vec