



Bag of Words

TOTAL POINTS 5

1. In this quiz, I would like you to find out about a particular type of feature used in machine learning. That feature is called "bag of words".

1 point

Use the internet to research the "bag of words" Representation and then answer the following questions. A simple web search should reveal some useful information. If you don't know the answer to any question, feel free to go back and do some more research before finished the quiz.

Firstly, what type of data does "bag of words" represent?

- ☐ Speech audio
- ☐ Images
- ☒ Written Text
- ☐ Music

2. The bag of words representation ignores the order of the words in a text

1 point

- ☒ True
- ☐ False

3. The bag of words representation ignores the number of times a word appears in a text

1 point

- ☐ True
- ☒ False

4. How many numbers are used to represent a piece of text

1 point

- ☐ One for every character (letter) in the language
- ☒ One for every unique word (i.e. not including repeats)
- ☐ One for every word in a text (the same as the word count)

5. What do you think are the advantages and disadvantages of bag of words?

1 point

Advantages:

1. Very simple to understand and implement.

Disadvantages:

1. Bag of words leads to a high dimensional feature vector due to the large size of Vocabulary, V.
2. Bag of words doesn't leverage co-occurrence statistics between words. In other words, it assumes all words are independent of each other.
3. It leads to a highly sparse vector as there is non-zero value in dimensions corresponding to words that occur in the sentence.



Your answer cannot be more than 10000 characters.



I, **Piyush Sambhi**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

[Learn more about Coursera's Honor Code](#)



Save

Submit