



**Congratulations! You passed!**

TO PASS 80% or higher

Keep Learning

GRADE

100%

## 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 / 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



**Correct**

Yes, that is right

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

1 / 1 point

- ☒ True
- ☐ False



**Correct**

Yes, one problem with bag-of-words is that it cannot gain information from word order

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

1 / 1 point

- ☐ True
- ☒ False



**Correct**

Yes, the representation is based on the number of times a word appears

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

1 / 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)



**Correct**

Yes, the numbers represent how many times a word is used. The basic bag of words includes a number for each word in a given text, but if you want to use it in machine learning you need the same number of features for every text, which means you need a number for each word in the language.

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

1 / 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.



**Correct**

Some advantages:

- Simple, clear representation
- The same number of features for each text
- Words are very important to meaning, so it is an important representation
- Words that appear a lot in a text have more importance

Disadvantages:

- The main disadvantage is that it does not take account of word order, so it loses important aspects of meaning
- It is a large representation that includes a lot of features (one for every word), most of which will be zero for a given text
- It can't take account of similarity between different words (word embeddings is a solution to this)