Consider the document:
"Information retrieval is the task of finding the documents satisfying the information needs of the user"

Using MLE to estimate the unigram probability model, what is $P(\text{the}|M_d)$ and $P(\text{information}|M_d)$?

1. 1/16 and 1/16
2. 1/12 and 1/12
3. 1/4 and 1/8
4. 1/3 and 1/6

Consider the following document

d = “information retrieval and search”

1. P(information search | M_d) > P(information | M_d)
2. P(information search | M_d) = P(information | M_d)
3. P(information search | M_d) < P(information | M_d)

In vector space retrieval each row of the matrix M corresponds to

A. A document
B. A concept
C. A query
D. A term

Applying SVD to a term-document matrix M. Each concept is represented in K

A. as a singular value

B. as a linear combination of terms of the vocabulary

C. as a linear combination of documents in the document collection

D. as a least squares approximation of the matrix M

The number of term vectors in the matrix K_s used for LSI

A. Is smaller than the number of rows in the matrix M

B. Is the same as the number of rows in the matrix M

C. Is larger than the number of rows in the matrix M

A query transformed into the concept space for LSI has …

A. s components (number of singular values)
B. m components (size of vocabulary)
C. n components (number of documents)