ECS763U/ECS763P - Natural Language Processing - 2022/23 - Semester 1

Started on	Friday, 23 December 2022, 2:23 PM
State	Finished
Completed on	Friday, 23 December 2022, 2:48 PM
Time taken	24 mins 43 secs
Grade	**9.67** out of 10.00 (**97**%)

Quiz navigation

1 2 3 4 5 6 7 8 9 10

Show one page at a time

Finish review

Question **1**
Correct
Mark 1.00 out of 1.00
Flag question

The purpose of using a weighting technique like TF-IDF is to minimise the effect of unimportant words which appear in _____.

Select one:
- a. many documents ✓
- b. few documents

The correct answer is: many documents

Question **2**
Correct
Mark 1.00 out of 1.00
Flag question

Word2Vec is a technique like term-context matrix building which uses defined contexts of words as the basis for its vectors.

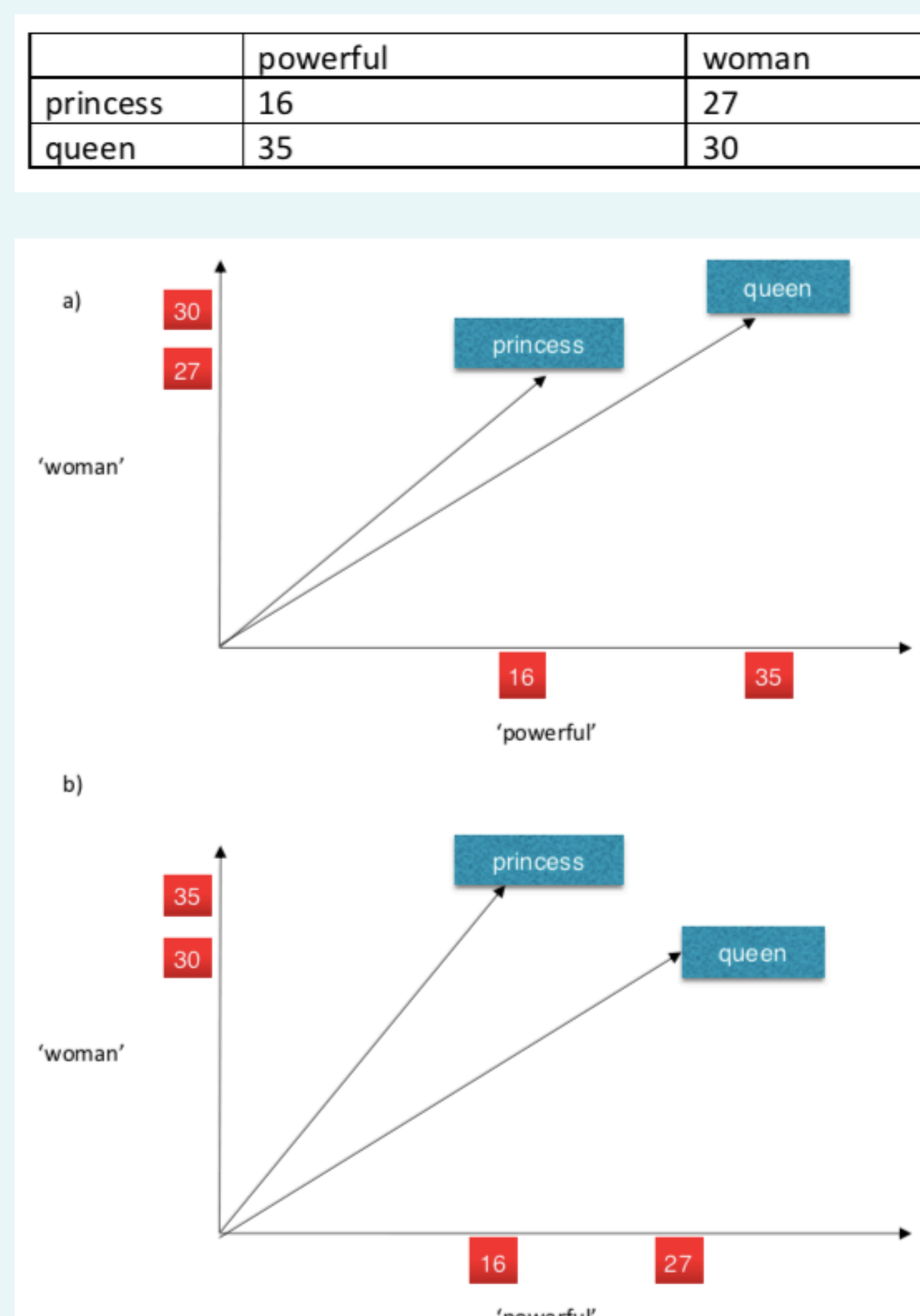
Select one:
- True
- False ✓

The correct answer is 'False'.

Question **3**
Correct
Mark 1.00 out of 1.00
Flag question

See the term-context matrix for two target words `princess' and `queen' below. Select the correct depiction of the two vectors representing the two words in 2-dimensional space from the two plots.

	powerful	woman
princess	16	27
queen	35	30



Select one:
- a. a ✓
- b. b

The correct answer is: a

Question **4**
Correct
Mark 1.00 out of 1.00
Flag question

Which of the following structure preserving maps have been experimented with between a syntactic system and vector space semantics system?

Select one or more:
- a. Lambek calculus -> vector spaces. ✓
- b. pregroup grammar -> vector spaces. ✓
- c. Domain of individuals -> vector spaces.
- d. Properties -> vector spaces.
- e. CCG -> vector spaces. ✓
- f. Referents -> vector spaces.

The correct answers are: pregroup grammar -> vector spaces., Lambek calculus -> vector spaces., CCG -> vector spaces.

Question **5**
Partially correct
Mark 0.67 out of 1.00
Flag question

Select all of the statements true of TF-IDF.

Select one or more:
- a. The IDF part of TF-IDF is a function of document frequency of a term divided by the number of documents. ✗
- b. TF-IDF gives more importance to words that occur in few documents. ✓
- c. TF-IDF is always guaranteed to give more realistic similarity scores.
- d. The IDF part of TF-IDF is a function of the number of documents divided by the document frequency of a term. ✓
- e. TF-IDF uses smoothing to avoid getting the logarithm of 0. ✓
- f. TF-IDF gives more importance to words that occur in many documents.

The correct answers are: TF-IDF uses smoothing to avoid getting the logarithm of 0., The IDF part of TF-IDF is a function of the number of documents divided by the document frequency of a term., TF-IDF gives more importance to words that occur in few documents.

Question **6**
Correct
Mark 1.00 out of 1.00
Flag question

Select all of the following statements which are consistent with the distributional hypothesis of language meaning.

Select one or more:
- a. You shall know a word by the company it keeps. ✓
- b. The meaning of a sentence is determined by its denotation according to a formal model of the world.
- c. Words that appear in the same contexts can be assumed to be synonymous. ✓
- d. Meaning is determined by a logical form attached to each word or constituent.
- e. The syntax of a language determines the meaning of sentences.
- f. A word's meaning can be approximated by the distribution of the contexts it occurs in. ✓

The correct answers are: You shall know a word by the company it keeps., A word's meaning can be approximated by the distribution of the contexts it occurs in., Words that appear in the same contexts can be assumed to be synonymous.

Question **7**
Correct
Mark 1.00 out of 1.00
Flag question

You have a corpus of 50 documents. See the table below for the document frequency for a list of terms in the corpus. What is the logarithm of the inverse document frequency (IDF) (not the full TF-IDF calculation) of the term `Han', using a log base of 10, to 3 decimal places?

t	DF*t* (document frequency of term)
Luke	30
Leia	22
Han	18
soldier	45
trooper	24
ewok	20
and	50
it	50

N= 50

Select one:
- a. 2.778
- b. 0.444 ✓
- c. -0.444
- d. 0.360
- e. -1.123

The correct answer is: 0.444

Question **8**
Correct
Mark 1.00 out of 1.00
Flag question

Consider the table of document frequencies for the terms in the table across a corpus of 50 documents shown below, in addition to the term-document matrix shown for three of the 50 documents in the corpus. Using a log base of 10 and using add-one smoothing to term frequencies, calculate the TF-IDF score of the term `soldier' in the document `Star Wars Episode VI: Return of the Jedi'. Give your answer to 2 Decimal Places.

t	DF*t* (document frequency of term)
Luke	30
Leia	22
Han	18
soldier	45
trooper	24
ewok	20
and	50
it	50

N= 50

t	TF*t,d* (frequency of term *t* in document *d*)		
	Star Wars Episode IV: A New Hope	*Star Wars Episode V: The Empire Strikes Back*	*Star Wars Episode VI: Return of the Jedi*
Luke	43	50	39
Leia	38	30	36
Han	40	43	15
soldier	102	94	80
trooper	98	89	78
ewok	0	0	30
and	1009	897	988
it	754	659	680

Answer: 0.09 ✓

The correct answer is: 0.09

Question **9**
Correct
Mark 1.00 out of 1.00
Flag question

Consider the table of document frequencies for the terms in the table across a corpus of 50 documents shown below, in addition to the term-document matrix shown for three of the 50 documents in the corpus. Using a log base of 10 and using add-one smoothing to term frequencies, calculate the TF-IDF score of the term `ewok' in the document `Star Wars Episode VI: Return of the Jedi'. Give your answer to 2 Decimal Places.

t	DF*t* (document frequency of term)
Luke	30
Leia	22
Han	18
soldier	45
trooper	24
ewok	20
and	50
it	50

N= 50

t	TF*t,d* (frequency of term *t* in document *d*)		
	Star Wars Episode IV: A New Hope	*Star Wars Episode V: The Empire Strikes Back*	*Star Wars Episode VI: Return of the Jedi*
Luke	43	50	39
Leia	38	30	36
Han	40	43	15
soldier	102	94	80
trooper	98	89	78
ewok	0	0	30
and	1009	897	988
it	754	659	680

Answer: 0.59 ✓

The correct answer is: 0.59

Question **10**
Correct
Mark 1.00 out of 1.00
Flag question

Consider the table of document frequencies for the terms in the table across a corpus of 50 documents shown below, in addition to the term-document matrix shown for three of the 50 documents in the corpus. Using a log base of 10 and using add-one smoothing to term frequencies, calculate the TF-IDF score of the term `Han' in the document `Star Wars Episode VI: Return of the Jedi'. Give your answer to 2 Decimal Places.

t	DF*t* (document frequency of term)
Luke	30
Leia	22
Han	18
soldier	45
trooper	24
ewok	20
and	50
it	50

N= 50

t	TF*t,d* (frequency of term *t* in document *d*)		
	Star Wars Episode IV: A New Hope	*Star Wars Episode V: The Empire Strikes Back*	*Star Wars Episode VI: Return of the Jedi*
Luke	43	50	39
Leia	38	30	36
Han	40	43	15
soldier	102	94	80
trooper	98	89	78
ewok	0	0	30
and	1009	897	988
it	754	659	680

Answer: 0.53 ✓

The correct answer is: 0.53