Week 7 Practice Quiz

Question 1

Which of the following word pairs are paradigmatically related? Check all that apply.

- Car, vehicle
- Computer, laptop
- Car, drive
- Computer, keyboard

Question 2

Which of the following word pairs are syntagmatically related? Check all that apply.

- Car, drive
- Computer, keyboard
- Car, vehicle
- Computer, laptop

Question 3

Suppose the pseudo-document representations for the contexts of the terms A and B in the vector space model are given as follows:

d_A = (0.30, 0.20, 0.40, 0.05, 0.00, 0.05) d_B = (0.40, 0.10, 0.30, 0.00, 0.20, 0.00)

What is the EOWC similarity score?

- 0.26
- 0.40
- 0.22
- 0.32

Question 4

True or false? Adding IDF (invert document frequency) weighting to the EOWC similarity function will penalize common words in the corpus.

- True
- False

Question 5

True or false? The EOWC score is non-negative and cannot exceed 1.

- True
- False

Question 6

Which of the following is not the reason that NLP is difficult?

- Word-level ambiguity, such as the word "design" can be a noun or a verb
- Syntactic ambiguity, such as PP attachment
- Coreference resolution, such as reference of "he," "she," or "it"
- Difficulty of transforming non-text data into text data

Question 7

Which of the following statements about NLP is correct?

- For English, parsing POS-tagging with 100% accuracy is possible if computing resource and time is unlimited.
- Shallow NLP tends to be robust, while deep NLP doesn't scale up well.
- POS-tagging is considered to be deep NLP.
- Summarization and translation are considered to be shallow NLP.

Question 8

For sentiment analysis (opinion mining), which kind of text representation is generally used?

- String
- Words
- Words and syntactic structures
- Words, syntactic structures, entities, and relations

Question 9

Paradigmatic related words have:

- High context similarity
- High co-occurrence

Question 10

Which of the following help to improve EOWC? Check all that apply.

- Use sublinear transformation of term frequency
- Reward matching rare words and discount matching frequent ones

Week 7 Quiz

Question 1

True or false? A paradigmatic relation is a relation between two words that tend to *co-occur* with each other, while a syntagmatic relation is between two words that tend to occur in a *similar* context.

- False
- True

Question 2

In a collection of English news articles, which word do you expect to have a higher IDF?

- “learning”
- “the”

Question 3

Suppose the pseudo-document representations for the contexts of the terms A and B in the vector space model are given as follows:

d_A = (0.10, 0.50, 0.00, 0.40, 0.00, 0.00) d_B = (0.20, 0.40, 0.30, 0.00, 0.10, 0.00)

What is the EOWC similarity score?

- 0.22
- 1
- 0.02
- 0.20

Question 4

True or false? Syntactic analysis (parsing) is an easier task than lexical analysis (part-of-speech tagging).

- True
- False

Question 5

"A man saw a boy with a telescope." What kind of ambiguity does the sentence have?

- Word-level ambiguity
- Syntactic ambiguity

Question 6

In an online text mining application where response time is the key factor to consider, what kind of NLP features can be used? Check all that apply.

- POS-tagging
- Word tokenization
- Relation extraction
- Syntactic parsing

Question 7

True or false? Deeper NLP requires more human effort and usually is less accurate.

- True
- False

Question 8

True or false? Word-based representation is not powerful.

- True
- False

Question 9

Which of the following is correct about paradigmatic and syntagmatic words relations?

- Paradigmatic related words have high co-occurrence.
- Syntagmatic related words have high context similarity.
- Monday, Tuesday are words of paradigmatic relation.

Question 10

Why does EOWC not work well?

- It favors matching frequent terms.
- It favors matching rare words.
- It treats words unequally.

Week 8 Practice Quiz

Question 1

You are given a unigram language model θ distributed over a vocabulary set V composed of **only** 4 words: “the”, “machine”, “learning”, and “data”. The distribution of θ is given in the table below:

w	**P(w|θ)**
machine	0.1
learning	0.2
data	0.3
the	0.4

P(“machine learning”|θ)=

- 0.02
- 0.3
- 0.2
- 0.004

Question 2

Assume the same unigram language model as in Question 1. Then, P(“learning machine”|θ)=

- 0.02
- 0.3
- 0.2
- 0.004

Question 3

Assume the same unigram language model as in Question 1. Then, P(“learning machine learning”|θ)=

- 0.004
- 0.02
- 0.3
- 0.2

Question 4

True or false? A random variable X with P(X=1)=1 achieves the minimum possible entropy.

- True
- False

Question 5

True or false? The outcome of an unbiased coin is easier to predict than the outcome of a biased coin.

- False
- True

Question 6

Which of the following is not true?

- If $H(X) = H(Y)$, then X and Y follow the same distribution
- If $H(X|Y) = H(Y|X)$, then $H(X) = H(Y)$
- $I(X;Y) = I(Y;X)$

Week 8 Quiz

Question 1

You are given a unigram language model θ distributed over a vocabulary set V composed of **only** 4 words: "the", "global", "warming", and "effects". The distribution of θ is given in the table below:

w	**P(w|θ)**
the	0.3
global	0.2
warming	0.2
effects	X

What is X, i.e., P("effects"|θ) ?

- 0.3
- 0.2
- 0.1
- 0

Question 2

Assume you are given the same unigram language model as in Question 1. Which of the following is **not** true?

- P("global warming"|θ)>P("warming global"|θ)
- P("global warming"|θ)=0.04
- P("the global warming effects"|θ)<P("global warming effects"|θ)
- P("text mining"|θ)=0

Question 3

True or false? Let Xtext, Xmining, and Xthe be binary random variables associated with the words "text", "mining", and "the", respectively. Assume that the probabilities of the random variables are estimated based on a large corpus. Then we should expect H(Xtext|Xmining)>H(Xtext|Xthe).

- False
- True

Question 4

True or false? I(X;Y)=0 if and only if X and Y are independent.

- True
- False

Question 5

Let w be a word and Xw be a binary random variable that indicates whether w appears in a text document in the corpus. Assume that the probability P(Xw=1) is estimated by Count(w)/N, where Count(w) is the number of documents w appears in and N is the total number of documents in the corpus.

You are given that "the" is a very frequent word that appears in 99% of the documents and that "photon" is a very rare word that occurs in 1% of the documents. Which word has a higher entropy?

- Both words have the same entropy.
- "the"
- “photon"

Question 6

Let X be a binary random variable. Which of the following is **not** true?

- If P(X=1)=1, then H(X) = 1
- $H(X) \leq 1$
- If P(X=0)=1, then H(X) = 0
- If P(X=0)=1, then H(X) = 1

Question 7

True or false? An unbiased coin has a higher entropy than any biased coin.

- True
- False

Week 9 Practice Quiz

Question 1

You are given two unigram language models, θ1 and θ2, as defined in the table below:

w	P(w|θ1)	P(w|θ2)
the	0.4	0.05
of	0.4	0.05
technology	0.1	0.5
machine	0.1	0.4

Suppose we are using a mixture model for document clustering based on the two given unigram language models, θ1 and θ2, such that P(θ1)=0.3 and P(θ2)=0.7. To generate a document, first, one of the two language models is chosen according to P(θi), and then all the words in the document are generated based on the chosen language model.

The probability of generating a document composed only of the one word "technology" using the given mixture model is P("technology")=

- 0.3
- 0.38
- 0.58
- 0.7

Question 2

Assume the same given as in Question 1. What is the probability of generating a document composed only of the phrase "the technology", i.e., P("the technology")?

- 0.0295
- 0.3
- 0.1444
- 0.0589

Question 3

In mixture model, why do different components tend to assign high probability on different words?

- Because it gives a higher overall likelihood
- Because the model was initialized with components with high probability assigned to different words
- Because during training, when different components assign high probability to the same model, the training restarts

Question 4

Why it is good to have the "background" component? Check all that apply.

- To better filter topic words into other components
- To prevent overfitting
- To improve model likehood

Question 5

What type of words are usually assigned with high probability in the background component?

- "the", "he", "she", "is"
- "car", "cat", "catch"
- "computer", "information", "data"

Question 6

Which of the following about the EM algorithm is false?

- It can be trapped into a local optimal solution.
- It always increase the likelihood.
- It is generally considered a fast algorithm for optimizing likelihood.
- The result of the EM algorithm does not depend on the initialization.

Question 7

In EM, what does the E-step do?

- Predicts values of unseen (hidden) variables
- Given the predicted values of unseen data, maximizes the joint likelihood

Question 8

Which of the following generative descriptions is not TRUE about PLSA?

- To generate a document, a distribution of topic weights (multinomial distribution) is assumed, which is considered part of the model.
- To generate a word, a topic is drawn from the document's topic weight distribution, and a word is drawn according to the topic's word distribution.
- To generate a topic assignment for a word, a coin is tossed to decide if the topic is from the background topic or not, and the probability of the background is a constant specified by the user.

Question 9

True or false? Let $\theta 1,...,\theta k$ be the k unigram language model's output by PLSA. Then, for a specific word w, the following relation always holds: $\sum i=1kP(w|\theta i)=1$.

- False
- True

Week 9 Quiz

Question 1

You are given two unigram language models θ1 and θ2 as defined in the table below:

w	P(w|θ1)	P(w|θ2)
concert	0.1	0.4
music	0.1	0.4
data	0.4	0.1
software	0.4	0.1

Suppose we are using a mixture model for document clustering based on the two given unigram language models, θ1 and θ2, such that P(θ1)=0.5 and P(θ2)=0.5. To generate a document, first, one of the two language models is chosen according to P(θi), and then **all** the words in the document are generated based on the chosen language model. The probability of generating the document d: "music software" using the given mixture model is P("music software")=

- 0.04
- 0.05
- 0.5
- 0.6

Question 2

Assume the same unigram language models, θ1 and θ2, defined as in the table of Question 1 with P(θ1)=0.5 and P(θ2)=0.5. We now want to generate documents based on the mixture model used in topic modeling. To generate a document **for each word**, we first choose one of the two language models, θ1 and θ2, and then generate the word according to the chosen model. The probability of generating the document d: "music software" according to this mixture model is P("music software")=

- 0.0625
- 0.625
- 0.125
- 0.0125

Question 3

We want to run PLSA on a collection of N documents with a fixed number of topics k where the vocabulary size is M. What is the number of parameters that PLSA tries to estimate? Consider each P(w|θj) or πd,j as a separate parameter.

- Mk+Nk
- Mk
- Nk
- MNk

Question 4

You are given a document d that contains only two words: “the” and “machine”. Assume that this document was generated from a mixture of two unigram language models: a known background language model θB and an unknown topic language model θd. Let $P(\theta B)=\lambda$ and $P(\theta d)=1-\lambda$ and assume that $P(\text{“the”}|\theta B)=0.9$ and $P(\text{“machine”}|\theta B)=0.1$. We want to estimate θd using maximum likelihood. Then, as λ increases, $P(\text{“machine”}|\theta d)$ will:

Hint: First get the maximum likelihood estimates of the two words in θd (refer to the lecture on "Probabilistic Topic Models: Mixture Model Estimation"). Then, write $P(\text{“machine”}|\theta d)$ as a function of λ and study the behavior of the function.

- Increase
- Decrease
- Remain the same

Question 5

True or false? In general, PLSA using the EM algorithm does not stop until it achieves the global maximum of the likelihood function.

- False
- True

Question 6

True or false? Let $\theta 1,...,\theta k$ be the k unigram language model's output by PLSA and V be the vocabulary set. Then, for any $i\in\{1,...,k\}$, the following relation always holds: $\sum_{w\in V} P(w|\theta i)=1$.

- True
- False

Question 7

True or false? The EM algorithm **cannot** decrease the likelihood of the data.

- True
- False

Question 8

True or false? Assume that the likelihood function of PLSA has multiple local maxima and one global maximum. There exists an initial set of parameters for which PLSA will converge to the global maximum of the likelihood function.

- True
- False

Question 9

True or false? When using PLSA to mine topics from a text collection, the number of parameters of the PLSA model stays the same as we keep adding new documents into the text collection assuming that the new documents do not introduce new words that have not occurred in the current text collection.

- False
- True

Week 10 Practice Quiz

Question 1

Which of the following is NOT a use of text clustering?

- Grouping similar documents together
- Grouping similar words together
- Grouping similar websites together
- Grouping similar pictures together

Question 2

Which of the following is TRUE about the mixture model?

- Words of the document are drawn from a mixture of topics where the mixing weight depends on different documents.
- Topics are a mixture of words where the mixing weight depends not only on the topics but also the documents.

Question 3

Which of the following is NOT true about the maximal likelihood of a set of documents?

- If we exchange every word "A" and "B", the maximal likelihood does not change.
- If we have a document "w1 w2 ... wn" changed into "wn ... w2 w1", the maximal likelihood does not change.
- if we have every document doubles (a document "w1 w2 ... wn" becomes "w1 w1 w2 w2 ... wn wn"), then the maximal likelihood does not change.

Question 4

If we have a large collection of documents to train PLSA with, what is the best way to initialize the model?

- Randomly initialize
- Initialize each topic as a distribution with probability 1 on a random single word but zero everywhere else and documents' topic weight to be 1 on a random topic but 0 everywhere else
- Train PLSA on a small subset collection of documents and use the model to initialize, and for other documents randomly initialize the documents' topic weights

Question 5

Which of the following is correct about K-means and PLSA?

- Only the results of PLSA depend on the way it was initialized.
- Both algorithms require the user to specify the number of clusters/topics.
- Only K-means is an iterative algorithm.
- Both of them have a clear objective function.

Question 6

What is the disadvantage of using a model-based clustering algorithm?

- It is difficult to substitute a different similarity measure.
- It's much slower to train.
- The performance is much worse than other methods.

Question 7

What is the difference between direct and indirect evaluation for a clustering algorithm? Check all that apply.

- Direct evaluation requires a human annotated gold standard cluster.
- Indirect evaluation requires a user specified application to test with.
- Direct evaluation is better than indirect evaluation.

Week 10 Quiz

Question 1

What is NOT the motivation for text clustering?

- To quickly get an idea about a large collection of documents
- To link similar documents and remove duplicated documents
- To remove spam documents based on a small collection of human annotated spam documents
- To create structure of text data

Question 2

In the EM algorithm, which step improves the model likelihood?

- E-step
- M-step

Question 3

True or false? In the EM algorithm, the model likelihood monotonically increases.

- True
- False

Question 4

What is the most difficult part of directly applying maximal likelihood to PLSA?

- The objective function needs to sum over all topics for each word.
- The objective function needs to sum over all words for each document.
- The objective function needs to sum over all documents in the collection.

Question 5

For the agglomerative clustering algorithm, which of the following is not TRUE?

- It's a bottom-up algorithm to form a hierarchy.
- The depth of the hierarchy is always log2(N) where N is the number of items.
- The user needs to specify a similarity measurement.

Question 6

Which evaluation method is best for clustering results of a large collection of documents?

- Use the direct evaluation method and create human annotations for each document in the collection.
- Use the indirect evaluation method and test performance for an application with or without clustering.

Question 7

Which of the following is a generative classification algorithm?

- Naive Bayes
- K-NN
- Logistic Regression

Week 11 Practice Quiz

Question 1

Sentiment classification can be treated as a text categorization problem.

- True
- False

Question 2

Assume that documents are being classified into 3 categories, c1, c2, and c3 such that a document can belong to multiple categories. The table below shows the prediction of a classifier, denoted by "y" or "n", in addition to the true label (ground truth) represented by a "+" or "-", where a correct prediction is either y (+) or n (-).

	c1	**c2**	**c3**
D1	y(+)	y(-)	n(+)
D2	n(-)	y(+)	n(-)
D3	y(+)	n(-)	y(+)
D4	y(+)	y(+)	y(+)

Let P(ci), R(ci), and F(ci) denote the precision, recall, and F1 measure associated with category ci, respectively.

Which of the following is **not** true?

- P(c3) = 2/3 R(c3) = 1
- P(c1) = 1 R(c1) = 1
- F(c2) = F(c3) = 4/5

Question 3

Given the same data as in Question 3, what are the **precision** and **recall** values of the classifier using **micro-averaging** (i.e., by pooling all decisions together)?

- P = 7/8 R = 7/8
- P = 7/12 R = 7/12
- P = 7/12 R = 8/12
- P = 1 R = 1

Question 4

Which of the following is not true?

- Naive Bayes is a generative classifier while K-NN is discriminative.
- Logistic Regression try to estimate d+1 weights associated with d features.
- K-NN tries to estimate d+1 weights associated with d features.

Question 5

Suppose we have the following training dataset of emails where each email is associated with the label spam or ham (not-spam). We want to train a Naive Bayes classifier based on this dataset.

- d1 is Spam and have words: Save Money No Fees
- d2 is Ham and have words: Back to the Future
- d3 is Ham and have words: Back to School Night

Using maximum likelihood estimation without smoothing, what is P(Spam)?

- 1/3
- 1/2
- 1/4
- 1/5

Question 6

Assume the same given as in Question 6 and that additive probability smoothing is being used to evaluate P(w|Spam) and P(w|Ham), i.e., $P(w|Spam)=c(w,Spam)+1\sum w' \in Vc(w',Spam)+|V|$ and $P(w|Ham)=c(w,Ham)+1\sum w' \in Vc(w',Ham)+|V|$ where $|V|=10$ is the size of the vocabulary.

Which of the following documents has the **highest** probability of being classified as **spam** by the Naive Bayes classifier?

Hint: You should not need to compute the actual probabilities to answer this question. You can answer it by inspecting the score function on the slide entitled "Anatomy of Naïve Bayes Classifier." (Lecture 11.1)

- “No fees”
- “Save money back”
- “Save money future”
- “Future school no fees”

Question 7

To apply Naive Bayes classification, we first need to estimate the parameters $P(w|\theta i)$ and $P(\theta i)$ for each corresponding category i. Suppose we would like to do binary classification. Consider the following corpus of two documents, d1 and d2 associated with two categories, T1 and T2. Each category contains one document as follows:

T1:{d1=(w1w1w1w1w3w3)}

T2:{d2=(w1w1w2w2w3w4)}

We estimate the parameters using the maximum likelihood estimator, i.e., $P(w|\theta\_i)=c(w,T\_i)|T\_i|$ and $P(\theta\_i)=|T\_i|/\sum\_j|T\_j|$, where $|T\_i|$ is the total number of words in category i.

Given a new document d3=(w3,w4), what will $P(d3|\theta 1)$ be?

- 0
- 1
- 0.6
- 0.5

Question 8

Suppose that we now use Laplace smoothing, what is P(θ1|d3)?

Note that Laplace smoothing is an additive smoothing method that is defined by P(w|θi)=(c(w,Ti)+1)/(|Ti|+|V|) where |V| is the size of the vocabulary in the training data (i.e., the number of unique terms in the training data).

- 1/2
- 3/7
- 5/8
- 2/29

Question 9

Which category would Naive Bayes predict for d3 if we use Laplace smoothing?

- Category 1
- Category 2

Question 10

The following table shows the similarity values between a set of emails as well as a binary label associated with each email indicating whether it is spam (label=1) or ham (label=0).

	D1	D2	D3	D4	D5	D6	Label
D1	100	0.1	0.5	0.8	0.82	0.85	1
D2	0.1	1000	0.85	0.05	0.12	0.7	0
D3	0.5	0.85	10000	0.1	0.1	0.6	0
D4	0.8	0.05	0.5	100000	0.9	0.1	1
D5	0.82	0.12	0.1	0.9	1000000	0.3	1
D6	0.85	0.7	0.6	0.1	0.3	1.0	?

Suppose we use {D1,D2,D3,D4,D5} as our training dataset and use the k-Nearest Neighbor classifier to predict the label of email D6. If k=1, then the prediction of the classifier for D6 is:

- 1
- 0
- There is a tie and thus 0 or 1.

Question 11

Assume the same setup as in Question 11. If k = 2, then the prediction would be:

- 1
- 0
- There is a tie and thus 0 or 1.

Week 11 Quiz

Question 1

Assume that documents are being classified into two categories, c1 and c2, such that a document can belong to more than one category. The table below shows the prediction of a classifier, denoted by “y” or “n", in addition to the true label (ground truth) represented by a “+” or “-”, where a correct prediction is either y (+) or n (-).

	c1	**c2**
D1	y(+)	y(+)
D2	n(-)	y(+)
D3	n(+)	n(-)
D4	y(-)	y(+)
D5	n(+)	n(-)

Let P(ci) and R(ci) denote the precision and recall associated with category ci, respectively. The precision and recall of c1 and c2 are:

- P(c1) = 1/2 R(c1) = 1/2 P(c2) = 1 R(c2) = 1
- P(c1) = 1/3 R(c1) = 1/2 P(c2) = 1 R(c2) = 1
- P(c1) = 1/2 R(c1) = 1/3 P(c2) = 1 R(c2) = 1
- P(c1) = 1/2 R(c1) = 1/2 P(c2) = 1/2 R(c2) = 1/2

Question 2

Given the same data as in Question 1, the classification accuracy of the classifier is:

- 8/10
- 3/10
- 9/10
- 7/10

Question 3

Given the same data as in Question 1, what is the recall of the classifier using **micro-averaging** (i.e., by pooling all decisions together)?

- 4/5
- 1
- 5/6
- 2/3

Question 4

Which one of the following statements is **not** an opinion?

- PLSA is a mixture model.
- PLSA is the best method for a topic mining task.
- PLSA always performs similarly to LDA.

Question 5

True or false? Word unigrams are the best performing features for sentiment classification.

- False
- True

Question 6

True or false? Suppose we are using logistic regression for binary classification (i.e., k=2) where the number of features is M. Then, the number of parameters to be estimated is M+1.

- True
- False

Question 7

True or false? Assume we are using word n-grams as features to perform sentiment classification. Then, higher values of n will usually be **less** prone to overfitting (i.e., for higher values of n, the difference between training and testing accuracies will be smaller).

- False
- True

Question 8

Why is accuracy sometimes not good for classification evaluation? Check all that apply.

- Some decisions are more serious than others.
- For imbalanced dataset, high accuracy does not imply good performance.
- Computation of accuracy is difficult.

Question 9

If you want to put more emphasis on precision than recall, how should you adjust the value of β?

- Choose a high value of β
- Choose a low value of β

Week 12 Practice Quiz

Question 1

True or false? NetPLSA leverages the power of both the text and the network structure to mine topics.

- True
- False

Question 2

True or false? NetPLSA tries to smooth the topic transitions by forcing neighbor nodes in the network to have different topic coverage.

- False
- True

Question 3

Contextual Probabilistic Latent Semantic Analysis (CPLSA) can be applied to which of the following tasks?

- Discovering temporal trends of topics in text
- Revealing how the coverage of topics in different locations evolves over time
- All of the above

Question 4

Suppose we are interested in discovering topics whose coverage in Twitter has strong correlations with airline prices. Which method would be best suited for this task?

- Iterative Topic Modeling with Time Series Feedback
- PLSA
- LDA
- Contextual PLSA (CPLSA)

Question 5

Deep learning is a new topic emerging in machine learning. Suppose we are interested in knowing whether US researchers and those outside the US have different focuses when working on this topic. For this purpose, we can collect research publications with metadata such as the author names, their affiliations, and locations. Which of the following text mining techniques is most suitable for this task?

- Contextual PLSA (CPLSA)
- Iterative Topic Modeling with Time Series Supervision
- Text clustering

Question 6

What is the additional input other than text in casual topic mining?

- Time
- Link
- Anchor text

Question 7

To measure the causality between two series, which of the following is true?

- Granger is often used.
- The correlation of the two values at different time stamps is needed.
- The time the two series arrive at peaks or dips is needed.

Question 8

If one is interested in finding out the trending topics in different countries in Twitter with the location information provided, what kind of technique should be used?

- Iterative Topic Modeling with Time Series Feedback
- PLSA
- LDA
- Contextual PLSA (CPLSA)

Week 12 Quiz

Question 1

Examine the objective function of NetPLSA in the lecture entitled **Contextual Text Mining: Mining Topics with Social Network Context**. Increasing λ will:

- Make neighbor nodes have more similar topic coverage
- Make neighbor nodes have less similar topic coverage
- Not affect the topic coverage of neighbor nodes

Question 2

You are given an undirected citation network composed of papers {p1,...,pn} as nodes, where a link between papers pi and pj means that one of the papers cited the other. Suppose you want to use the given data to discover the topics (research areas) of the papers. Which of the following methods is expected to work best?

Hint: Papers that have a citation relationship are more likely to belong to the same research area.

- NetPLSA
- CPLSA
- PLSA
- Sentiment analysis

Question 3

You are given a collection of news articles along with their publishing dates and want to reveal which topics have attracted increasing attention in a certain time period. Which of the following methods is most suitable for this task?

- CPLSA
- NetPLSA
- Sentiment analysis

Question 4

Imagine a company is interested in understanding any factors related to their fluctuating sales of a new product in the past year. They collected the companion text data including the consumer reviews of the product from multiple websites with time stamps in the past year and hope to gain potential insights from such text data. Which of the following text mining techniques would you recommend to them?

- Iterative topic modeling with time series supervision
- Contextual PLSA (CPLSA)
- Text clustering

Question 5

The US government implemented a new health care policy in year 2010. Suppose the government is interested in understanding the impact of such a policy and how the policy has affected what people talk about in social media. For this purpose, we can collect social media text data such as forum posts and tweets with time stamps before 2010 and after 2010. Which of the following text mining techniques is most suitable for such a text mining task?

- Contextual PLSA (CPLSA)
- Iterative Topic Modeling with Time Series Supervision
- Text clustering

Question 6

Context can be used to (check all that apply):

- Partition text
- Annotate topics

Question 7

Which of the following statement of CPLSA is NOT correct?

- CPLSA is an extension of PLSA.
- It models the joint probability of text and context.
- The EM algorithm can be used for optimization.
- It enables contextual text mining.