

CS 124 Winter 2019 Practice Final Exam

Relevant policies:

- The exam will only contain multiple choice questions. You will answer them on a separately provided bubble sheet.
- In the exam you may use a computer on which anything can be downloaded, but you may use the web only to view the `cs124.stanford.edu` website for our class, to view the edX website for our class, and to view the Piazza website for our class. You may not use other information on the web.
- You are not allowed to write programs to check answers during the exam.
- You may use the calculator functions on your computer to compute values for functions such as cosine.

Please note that this practice final exam is **shorter** than the actual final exam and is meant to give you a sense of the types of questions to expect. The actual final exam will contain more questions and will take longer to complete.

Regular Expressions and Edit Distance

1. Which of the following regular expressions matches the string "CS124" but does not match the string "CS221"?
 - (a) `CS[12]2[14]`
 - (b) `CS\d{2-4}`
 - (c) `CS12\d`
 - (d) `CS\d21`
2. Suppose we modify our edit distance formula such that a substitution has a cost of 3, and deletions and insertions each have a cost of 2. What is the number in the highlighted cell?

L					
A					
N					
I					
F					
#					
	#	E	X	A	M

- (a) 13
- (b) 10
- (c) 8
- (d) 6

Language Modeling and Naive Bayes

3. What is the **naive** assumption made by the Naive Bayes model used for classifying documents?
- (a) Words in a document occur independently of each other, given the class.
 - (b) The training data does not need to be labeled to achieve strong accuracy.
 - (c) The frequency of the classes in the given documents is not relevant for classification.
 - (d) Although gradient descent is probabilistic, it will converge after enough iterations.

4. Suppose we build a **unigram** language model from the following snippet of text.

<s> the cat wore the hat </s>
<s> the hat was on the cat </s>
<s> i want a blue hat the cat said </s>
<s> but alas the cat's hat was red </s>

Suppose we see the word 'the', what is the probability that the next word is 'hat', i.e. $P(\text{hat}|\text{the})$?

- (a) $\frac{2}{17}$
 - (b) $\frac{3}{51}$
 - (c) $\frac{1}{3}$
 - (d) $\frac{4}{17}$
5. You evaluate your Naive Bayes classifier on a dataset with 200 positive ($y = 1$) examples and 200 negative ($y = 0$) examples. It correctly classifies 160 examples as positive and classifies the remaining 240 examples as negative.

What is the recall?

- (a) 0.2
 - (b) 0.4
 - (c) 0.6
 - (d) 0.8
6. Given the following corpus, which of the following sentences is assigned the **highest probability** when using a bigram language model and **MLE estimates** (no Laplace smoothing). Don't forget to use start and end tokens.

Jess loves drinking coffee
Bill really enjoys coffee
Jess enjoys drinking coffee with Bill
Sal likes drinking with friends

- (a) Bill really loves coffee
- (b) Jess enjoys coffee
- (c) Jess loves drinking with Bill
- (d) Bill loves Jess

.

Logistic Regression and Sentiment Analysis

7. What role does the sigmoid function play in logistic regression?
- (a) It turns a probability value into a boolean value
 - (b) It squashes the output value $wx + b$ into a probability
 - (c) It converts the positive classes into positive values and negative classes into negative values, which helps with the classification step
 - (d) It doesn't play a role: it's used only in Naive Bayes

8. Suppose you are creating a logistic regression classifier to determine if a course review is positive or negative. Below is your training corpus.

I enjoyed this course! The material was interesting, the professor's lectures were engaging.

(positive)

dont do it **(negative)**

... just no ... **(negative)**

it is a very hard class so i would say do not take it ever **(negative)**

absolutely wonderful **(positive)**

Which of the following features would be best to include in your classifier to maximize performance on this corpus?

- (a) log of number of words in the review
 - (b) number of personal pronouns (I, you, etc.) in review
 - (c) average length of words in review
 - (d) number of punctuation marks in review
9. We are given the following bicycle review and we have to use logistic regression (using the sigmoid function) to classify it as either having positive sentiment ($y = 1$) or negative sentiment ($y = 0$).

*"This bicycle is really **shaky** and **uncomfortable**. Yet, it is such a **thrill** to ride. It reminds me of a **fun** rollercoaster. It is **great**, I **recommend** it!"*

Positive lexicon: thrill, fun, great, recommend

Negative lexicon: shaky, uncomfortable

Suppose $W = [1.7, -2.6, 0.4]$ and $b = 0.1$, use the following three features to calculate $P(y = 1|X)$ and $P(y = 0|X)$.

X_1 = count of positive lexicon in doc

X_2 = count of negative lexicon in doc

$X_3 = \ln(\text{word count of doc})$

Do not round off your numbers until the final solution.

(a) $P(y = 1|X) = 0.95, P(y = 0|X) = 0.05$

(b) $P(y = 1|X) = 0.05, P(y = 0|X) = 0.95$

(c) $P(y = 1|X) = 0.62, P(y = 0|X) = 0.38$

(d) $P(y = 1|X) = 0.38, P(y = 0|X) = 0.62$

Information Retrieval

10. Say we have the following four documents:

- I. the library over there has lots of books
- II. i have lots of books from the library
- III. there are lots of birds over there
- IV. i forgot to check if the library there is open

Which pair of documents has the highest Jaccard similarity?

- (a) I and II
- (b) I and III
- (c) II and IV
- (d) III and IV

11. What is the mean average precision for the following three sequences of retrieved documents, where R represents a relevant document and N represents a non-relevant document?

Sequence 1: R, N, R, R, N

Sequence 2: N, N, R, R, R

Sequence 3: R, N, R, N, N

- (a) 0.533
- (b) 0.439
- (c) 0.561
- (d) 0.706

Note: Use the following Table for the next three questions.

Use the following term frequencies (raw counts) for a few words in a collection of 4 documents. Use tf-idf weighting (LTC) and assume that these are the only documents and words in the collection.

term	Doc1	Doc2	Doc3	Doc4
bicycle	36	30	0	3
tree	25	0	6	0
book	8	10	20	17
computer	1	38	35	0

12. Which of the following is the unnormalized tf-idf vector for Doc1?

- (a) [0.32, 0.722, 0.0, 0.125]
- (b) [0.31, 0.0, 0.0, 0.323]
- (c) [0.185, 0.0, 0.0, 0.0]
- (d) [2.556, 2.398, 1.903, 1.0]

13. Which of the following is the unnormalized tf-idf vector for Doc3?

- (a) [0.31, 0.0, 0.0, 0.323]
- (b) [0.0, 0.535, 0.0, 0.318]
- (c) [0.185, 0.0, 0.0, 0.0]
- (d) [0.0, 1.778, 2.301, 2.544]

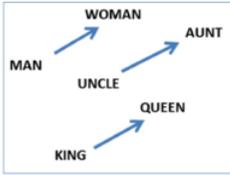
14. What is the cosine similarity between Doc1 and Doc3?

- (a) 0.856
- (b) 0.426
- (c) 0.390
- (d) 0.678

Relation Extraction and Vector Semantics

15. Given the vectors below, what is the value of X in the following equation?

$$\text{vector}(\text{'uncle'}) - \text{vector}(\text{'man'}) + \text{vector}(\text{'woman'}) = \text{vector}(X)$$



- (a) Queen
(b) Aunt
(c) King
(d) None of the above
16. How would you represent the following statement using Resource Description Framework (RDF) triples?

Stanford is a private university.

- (a) Stanford, type, private university
(b) organisation/type/private
(c) stanford/type/private
(d) Stanford IS-A private university
17. Suppose we run the Dipre algorithm on the following text instances starting with the following seeds.

Seeds:

(Samoyed, dog)

(Ragdoll, cat)

Text:

One kind of cat is a Ragdoll

A Samoyed is a dog breed

The Samoyed, a dog, is very friendly

Which of the following patterns will we have extracted after one iteration? Assume that the first entry in the seed tuple is replaced by $?x$ and the second entry is replaced by $?y$ with words and punctuation marks separated by spaces. Choose a single answer.

- (a) $?x$ is a $?y$

- (b) The $?x$, a $?y$, is very friendly
 - (c) $?x$, a $?y$
 - (d) One kind of $?x$ is a $?y$
18. Suppose we have the embedding for a word a , and we find another word b that has an embedding with a very high cosine similarity. Which of the following is most likely to be true?
- (a) Words a and b are synonyms.
 - (b) The embeddings for a and b are only alike due to randomness in the vector space.
 - (c) Words a and b appear in a lot of the same contexts.
 - (d) Words a and b can be used interchangeably without affecting the meaning of sentences.
19. Suppose we run the SNOWBALL algorithm on the text below to attempt to extract the FOUNDER-OF relation. Which of the patterns below will extract at least one correct example of that relation without extracting any incorrect ones (select all that apply)?

ORG entities are **bold**, and PERSON entities are underlined.

Correct examples:

- (**Microsoft**, Bill Gates)
- (**Facebook**, Mark Zuckerberg)
- (**Google**, Larry Page)
- (**Google**, Sergey Brin)

Source text:

Microsoft, founded by Bill Gates, produces both computer software and personal computers. The founders of **Google**, Larry Page and Sergey Brin, developed an advanced search experience. And Mark Zuckerberg, founder of **Facebook**, crafted a new communication platform. And, usage exists between them: indeed, Bill Gates is a user of **Google** search, and Larry Page of **Microsoft** products such as Word. Bill Gates of **Microsoft**, Larry Page and Sergey Brin of **Google**, and Mark Zuckerberg of **Facebook** were all pioneers of today's technology.

You can assume that all of the patterns are well formed SNOWBALL patterns.

- (a) **ORG**, founded by PERSON
- (b) **ORG**, PERSON
- (c) founders of **ORG**, PERSON
- (d) PERSON of **ORG**

Question Answering and Chatbots

20. Which of the following chatbots was the first to pass the Turing test?

- (a) Eliza
- (b) Perry
- (c) Microsoft Tay
- (d) Woebot

21. Consider “IRMA v1”, an IR-based question-answering system with the following interesting property: when it was last tested on a set of questions, IRMA’s accuracy was the exact same as its mean reciprocal rank. (Both were less than 1.)

Which of the following “improvements” to IRMA v1 is guaranteed NOT to improve its accuracy (on the same set of questions), if it was the only thing changed?

- (a) Retrieving better passages
- (b) Extracting answer candidates from the passages better
- (c) Ranking the candidate answers better

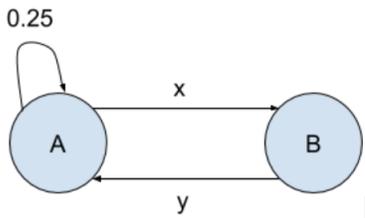
Hint: It may help to consider a concrete value for the accuracy and mean reciprocal rank, such as $\frac{1}{2}$. What does it mean to have a mean reciprocal rank of $\frac{1}{2}$, if your accuracy is also $\frac{1}{2}$?

Recommender Systems

22. True or False: TF-IDF can be used to pick important features for item profiles for content-based recommender systems.
- (a) True
 - (b) False
23. True or False: For the same recommendation task, it is possible to use the same utility matrix for both user-user collaborative filtering and item-item collaborative filtering.
- (a) True
 - (b) False
24. True or False: In collaborative filtering, the user vectors always have the same dimensions as the item vectors.
- (a) True
 - (b) False
25. At Stanford, students rate the quality of instruction of a course on a scale from 1-5 at the end of the quarter. As Carta's newest recommendation engineer, you have decided to binarize the course recommendation system with a threshold of 2.5 and use item-item collaborative filtering to recommend classes. Which of the following are valid observations about this approach? Assume that the following statements are true: most student reviews are 3, 4, or 5; students can opt out of providing a course rating.
- I. Since most student ratings are 3, 4, or 5, binarizing the ratings with a threshold of 2.5 causes loss of useful information about finer-grained user ratings data.
 - II. A new course faces no adverse effects during the period where no students have rated it yet.
 - III. It is more difficult for a low-enrollment course to overcome the cold start effect than a high-enrollment course.
 - IV. Students can opt out of providing a course rating, so the ratings provided to the recommendation system represent a biased sample of the population.
- (a) I, II, III, and IV
 - (b) I, III, and IV
 - (c) I and II only
 - (d) II, III, and IV

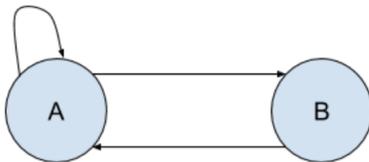
Networks

26. Suppose we have the following Markov chain. What are the transition probabilities, x and y ?



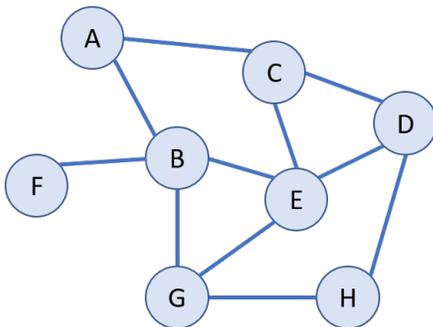
- (a) $x = 0.25, y = 0.75$
- (b) $x = 0.75, y = 0.25$
- (c) $x = 0.25, y = 1$
- (d) $x = 0.75, y = 1$

27. True or false: The following Markov chain is ergodic.



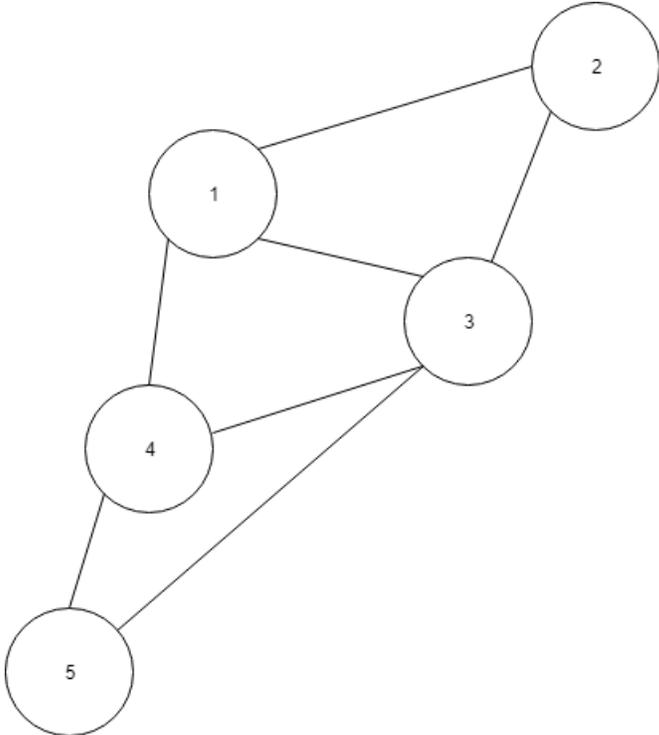
- (a) True
- (b) False

28. How many local bridges are in the following graph?



- (a) 3
- (b) 5
- (c) 7
- (d) 9

29. Given the following 5 node graph, which node has the highest betweenness centrality?



- (a) 1
- (b) 2
- (c) 3
- (d) 4