

CSCI 5832 Spring 2009 Quiz 1

Name: _____

On my honor, as a University of Colorado at Boulder student, I have neither given nor received unauthorized assistance on this work. _____.

1. **(5 points)** What is your favorite color? **Blue**
2. **(5 points)** Construct an FSA to recognize the set of strings over the alphabet a, b such that there are no consecutive repeated characters.

There are lots of answers to this. See the attached sheet.

3. **(5 Points)** Give a regular expression that corresponds to your machine (not to the language).

Depends on your answer

4. **(5 Points)** What is meant by the terms regular/irregular in the context of inflectional morphology?

Regular words are those that followed the normal inflectional rules given their word class (e.g., -ed rule for past tense verbs or -s for plural nouns). Irregular words don't.

5. **(5 points) True or False:** Regular relations are closed under intersection. False.
6. Consider the following Dr. Seuss rhyme. See extra sheet.

One fish two fish red fish blue fish black fish blue fish

- a) **(5 points)** Show a table with the bigram counts (as in the text) for this corpus. Given this table, give the **P(fish|two)** and **P(black|fish)**.
 - b) **(5 points)** Compute the probability mass that Good-Turing would assign to zero count bigrams given this corpus.
7. **(5 Points)** What is meant by the terms *Open Class* and *Closed Class* in the context of lexical categories? Give examples of each in your answer.

Open class terms are members of lexical classes that admit new members (i.e., nouns, verbs, etc.); closed class terms are members of lexical classes that don't (i.e., prepositions, determiners, etc.)

8. (5 Points) Give the mathematical equation that summarizes the generic statistical sequence labeling approach to part of speech tagging.

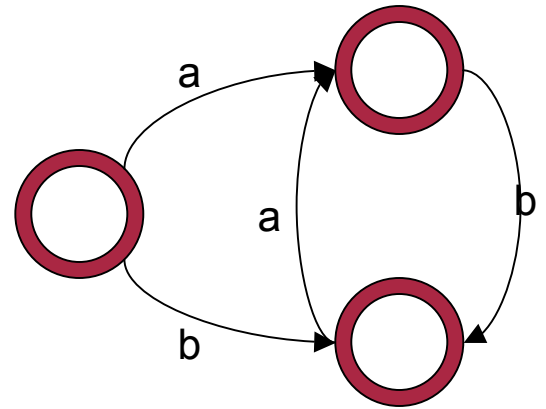
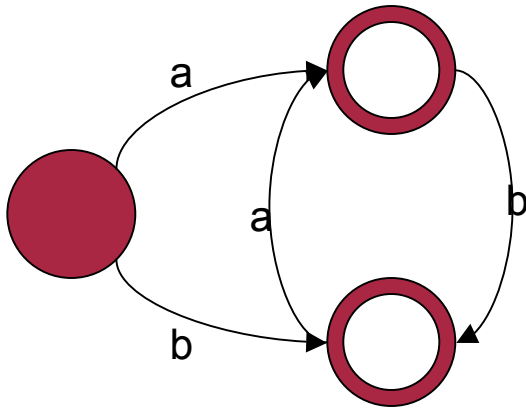
$$\operatorname{argmax} P(\textit{Tags}|\textit{Words})$$

9. (10 points) True or False: The Viterbi algorithm computes the probability of a sequence of observations given a hidden Markov model.

False. Viterbi finds the best state sequence given an observation sequence.

Question 2

There are lots of ways to do this. But the key is to decide what you want the machine to accept, not to model the rejects.



Question 3

$a(ba)^*b?$ | $b(ab)^*a?$

Question 5

	One	Fish	Two	Red	Blue	Black
One		1				
Fish			1	1	2	1
Two		1				
Red		1				
Blue		2				
Black		1				

a)

$$P(\text{fish}|\text{two}) = \text{Count}(\text{"two fish"})/\text{Count}(\text{"two"}) = 1/1 = 1$$

$$P(\text{black}|\text{fish}) = \text{Count}(\text{"fish black"})/\text{Count}(\text{"fish"}) = 1/6 = .166666$$

b)

$$N_0 = N_1/N \quad 7/11$$