

# Practice Quiz 1 -- Fall 22

Started: Sep 21 at 12:56pm

## Quiz Instructions

Practice Quiz 1, Chapters 2, 3, 4, 5, and 8.

### Question 1

1 pts

If you initialize byte-pair encoding with a vocabulary of size of 9 and then run the algorithm for 8 iterations, how large is the resulting vocabulary?

### Question 2

1 pts

Which of the following statements follows from Heap's Law when applied with the usual values for its constants.

- ☐ b. The size of a vocabulary grows without bound as a function of corpus size.
- ☒ a. Growth in the size of a vocabulary approaches a fixed ceiling as a function of the size of the corpus.

### Question 3

5 pts

Lemmatization and stemming are both methods for reducing surface word forms to underlying forms for some computational or linguistic purpose.

☒ True

☐ False

<s> some are fast </s>  
<s> and some are slow </s>  
<s> some are high </s>  
<s> and some are low </s>

#### Question 4

1 pts

Using the corpus given above, what is the conditional probability of seeing "some" having just seen the word "and" (ie.  $P(\text{some} \mid \text{and})$ .) Assume a bigram approach and no smoothing. Treat <s> and </s> as regular tokens (i.e., include them in the counts).

☐ 1

☐ 0.25

☐ .5

☐ 0

**Question 5****1 pts**

Given the earlier corpus, and a bigram language model (with no smoothing) what is the probability that will be assigned to the following sentence? Give your answer in a numerical form (not a fraction).

<s> some are low </s>

**Question 6****1 pts**

Assume a 2-class document classification setting (e.g. spam detection) where you are provided with a training set (positive and negative examples) and you are using naive Bayes.

True or False: In this application, the size of the vocabulary  $|V|$  is based on the number of word types in the corpus as a whole (both positive and negative examples).

---

☐ True

---

☐ False

**Question 7****1 pts**

Assume you have a 3-way authorship classification problem that you are addressing with logistic regression. Let's call the classes John, James, and Alex. During training you encounter a training document from class **James** with the feature vector  $[-3, 1, 4,$

1]. With the current set of weights, the model returns the softmax vector [0.2, 0.4, 0.4] over the classes John, James, and Alex respectively.

Give the gradient (vector of partial derivatives of the loss) for the **James** class that would be generated during training for this example using cross-entropy loss. Express

your answer as a vector of values as in [ , , ,  ]

### Question 8

1 pts

Which of the following best describes why English part-of-speech tagging is hard?

- ☐ Because the most frequently occurring words in English have multiple parts of speech.
- ☐ Because many English words have multiple parts of speech.
- ☐ Because English words can have many meanings.
- ☐ Because English spelling rules are so irregular.

A Matrix

From/To	Q1	Q2
Q1	0.6	0.4
Q2	0.5	0.5

B Matrix

--	--	--	--

	a	b	c
Q1	0.2	0.4	0.4
Q2	0.5	0.4	0.1

$\pi_i$

Q1	0.8
Q2	0.2

### Question 9

1 pts

Consider the preceding HMM setup, and an observation sequence of length 2 consisting of "c c".

Complete the final column of the Viterbi table given below

Time	1	2
Q1	0.32	<input type="text"/>
Q2	0.02	<input type="text"/>
Observation	c	c

No new data to save. Last checked at 2:14pm

Submit Quiz