Natural Language Processing

Assignment- 3

TYPE OF QUESTION: MCQ

Number of questions: 7 **Total mark: 10 (Q5, Q6, Q7 carries two marks each)**

Question 1:

Which of the following words contains both derivational and inflectional suffixes?

1. happiness
2. quicker
3. enjoyment
4. responsibilities

Answer: 4

Solution:
Responsibilities = respons(e) (Root word) + ible (derivational suffix) + ity (derivational suffix) + es (inflectional suffix).

Question 2:

Let's assume the probability of flipping heads two times in a row with a fair coin is q. Consider a sentence consisting of M random binary digits (0s and 1s). A model assigns probability to each digit in the sentence using the probability q. What is the perplexity of the sentence?

1. 2
2. 4
3. 8
4. 16

Answer: 2

Solution: The probability of flipping heads two times in a row is q=1/2×1/2=¼
Then perplexity is ((1/4)^M)^-1/M = 4

Question 3:
Assume that "x" represents the input and "y" represents the tag/label. Which of the following mappings are correct?

1. Generative Models - learn Joint Probability p(x, y)
2. Discriminative Models - learn Joint Probability p(x, y)
3. Generative Models - learn Posterior Probability p(y | x) directly
4. Discriminative Models - learn Posterior Probability p(y | x) directly

Answer: 1, 4
Solution: Generative classifiers learn a model of the joint probability p(x, y) and make their predictions by using Bayes rules to calculate p(y | x). Discriminative classifiers model the posterior p(y | x) directly, or learn a direct map from inputs x to the class labels y.

Question 4.

Natural language processing is essentially the study of the meaning of the words a human says or writes. Natural language processing is all around us all the time, but it also happens to be a way to improve the chatbot or product we interact with on a regular basis. Natural language processing is all about mimicking our own language patterns. Natural language processing can also improve the efficiency of business transactions and customer care. Natural language processing is the application of computer technology.

Suppose we want to check the probabilities of the *final words* that succeed the *string* language processing in the above paragraph. Assume d= 0; it is also given that no of unigrams = 78, no of bigrams = 122, no of trigrams = 130,, Question 6 and Question 7 are related to Question 5 corpus.

Solve the question with the help of **Kneser-Ney backoff technique.**

What is the continuation probability of *"is"* ?

1. 0.0078
2. 0.0076
3. 0.0307
4. 0.0081

Answer: 2

Solution: Refer week 3 lecture 12

Continuation probability of is = 1/130 = 0.0076

The numerator means the number of different string types preceding the final word, (here only 1 type– language processing is) and the denominator means the number of different possible n-gram types , in this case trigram = 130

Question 5:
What will be the value of P(is| language processing) using Kneser-Ney backoff technique and choose the correct answer below. . Please follow the paragraph in Question .

1. 0.5
2. 0.6
3. 0.8
4. 0.7

Answer: 3

Solution: Refer week 3 lecture 12

P(is| language processing) = ⅘ + 0*0.0076 = 0.8 [as d=0 so lambda = 0]

In this example is equal to the frequency of language processing *: the frequency of language processing is (here it occurs 4 times) plus the frequency of language processing can (occurs only once). Therefore, for word is, firstTerm(is) = 4/(4+1) = 0.8

Question 6.
What is the value of P(can| language processing)? Please follow the paragraph in Question 5

1. 0.1
2. 0.02
3. 0.3
4. 0.2

Answer: 4

Solution: Refer week 3 lecture 12

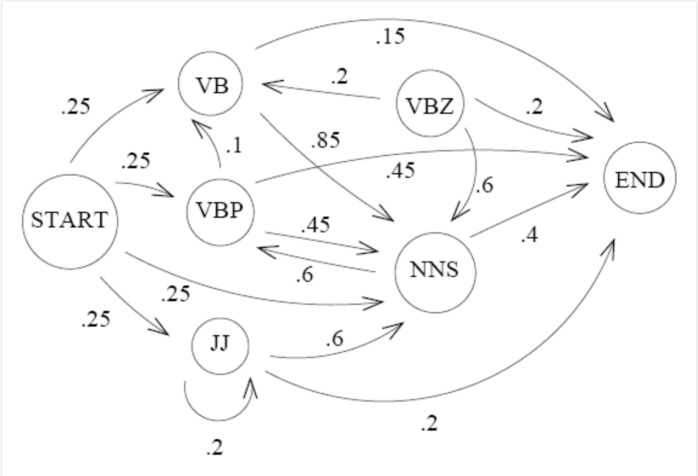
Similarly P(can| language processing) = ⅕ + 0*Continuation probability = 0.2

Language processing * occurs 5 times, language processing can occur only once.

Question 7:

Consider the HMM given below to solve the sequence labeling problem of POS tagging. With that HMM, calculate the probability that the sequence of words "free workers" will be assigned the following parts of speech;

VB NNS



	free	*workers*
JJ	**0.00158**	**0**
NNS	**0**	**0.000475**
VB	**0.00123**	**0**
VBP	**0.00081**	**0**
VBZ	**0**	**0.00005**

The above table contains emission probability and the figure contains transition probability

1. $4.80 * 10^{-8}$
2. $9.80 * 10^{-8}$
3. $3.96 * 10^{-7}$
4. $4.96 * 10^{-8}$

Answer: 4

Solution:

P(free workers, VB NNS)

= P(VB|*start*) * P(*free*|VB) * P(NNS|VB) * P(*workers*|NNS)

* P(end|NNS)

= 0.25 * 0.00123 * 0.85 * 0.000475 * 0.4

= $4.96 * 10^{-8}$
