Natural Language Processing

Assignment- 3

TYPE OF QUESTION: MCQ

Number of questions: 10 **Total mark: 10 X 1 = 10**

Question 1: Which of the following words contains both derivational as well inflectional suffixes:

1. regularity
2. carefully
3. older
4. availabilities

Answer: 4
Solution: availabilities = avail(Root word) + able (derivational suffix) + ity (derivational suffix) + es (inflectional suffix).

Question 2: Let's assume the probability of rolling 1 two times in a row of a dice is p. Consider a sentence consisting of N random digits. A model assigns probability to each of the digit with the probability p. Find the perplexity of the sentence.

1. 10
2. 6
3. 36
4. 3

Answer - 3

Solution: Probability of rolling 1 two times in a row is(⅙)*(⅙) = (1/36) . Then perplexity is ((1/36)^N)^-1/N = 36

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: Which one of the following is an example of the Generative model?

1. Conditional Random Fields
2. Naive Bayes
3. Support Vector Machine
4. Logistic Regression

Answer- 2

Solution: Others model in the option are discriminative model

---

Question 5. 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.

Solution 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

Question 6: 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

Question 7. 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

Question 8: Which of the following morphological process is true for motor+hotel → motel?

1. Suppletion
2. Compounding
3. Blending
4. Clipping

Answer: 3

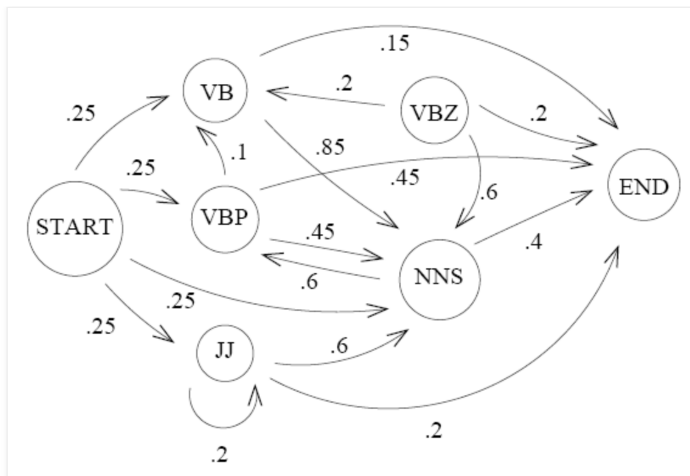
Solution:

Blending combines parts of two different words. Refer Lecture 13

Question 9: 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}

Question 10: Which of the following is/are true?

1. Only a few non-deterministic automation can be transformed into a deterministic one
2. Recognizing problem can be solved in linear time
3. Deterministic FSA might contain empty (ε) transition
4. There exist an algorithm to transform each automation into a unique equivalent automation with the least no of states

Answer: 2, 4

Solution:

Every non-deterministic automation can be transformed into a deterministic one. Deterministic FSA should not contain empty transition
Refer Lecture 14
