Q.1 Which of the following are true about information extraction?

A. It is the process of extracting meaningful information from text data.

B. It is the process of retrieving information from a corpus of text.

C. Both A and B

D. None of the above

Q2. Which of the following is/are a valid case of information extraction?

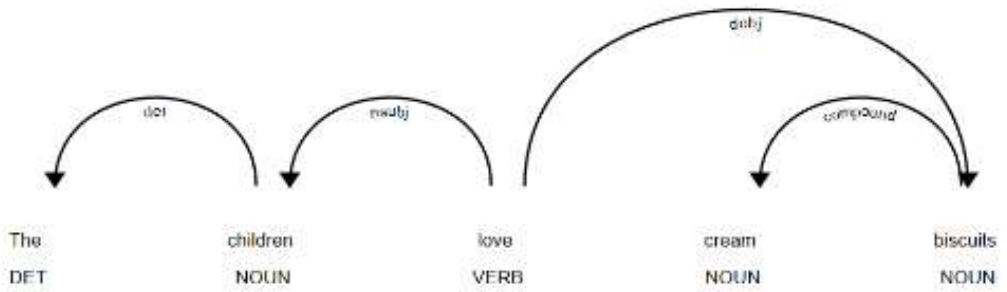
A. Looking for skill in resume

B. Extracting information from invoice

c. Finding relevant information from financial documents

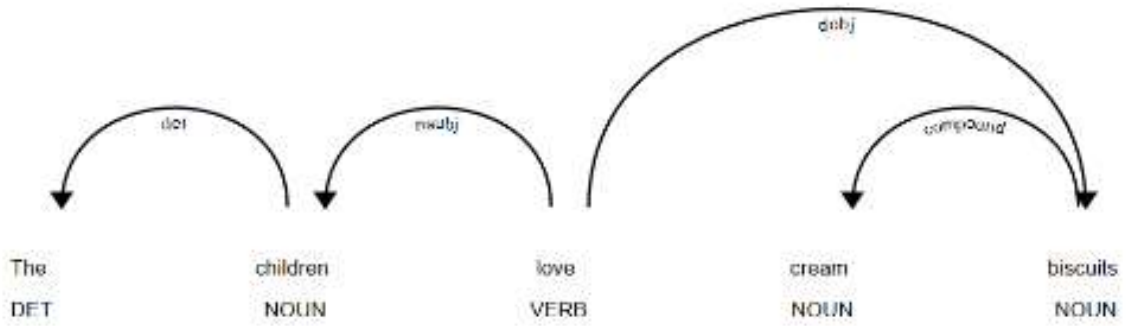
D. All of the above

Find the root word from the following:



A. Children
B. Love
C. Cream
D. Biscuits

Find the root word from the following:



A. Noun
B. Verb
C. DET
D. AUX

Q.5 Which of the following is/are true about dependency parsing?

A. It looks for dependency between all words in a sentence.

B. The word at the end of the arrow of a dependency graph is a child of the word at the arrowhead.

C. We can find how two words are related in a text.

D. All of the above.

Q.6 Find the named entity/entities from the following.

A. Google B. Happy C. Paris D. Play

Q7. Identify the named entities in the following sentence - "We never planned to head to Oxygen restaurant to have dinner with Charles."

A. restaurant

B. Oxygen

C. dinner

D. Charles

Q8. Which of the following is true?

A. Regex is a form of information extraction.

B. Only a rule-based approach can solve the information extraction problem.

C. Relation extraction is a form of information extraction.

D. All of the above.

Q9. The rule-based relation extraction approach requires working with ---- patterns.

A. Regex B. Hearst C. Supervised D. All of the above

Q10. Which one of the following formulas gives the normalized hamming distance?

A. Length of string/ Hamming distance

B. Hamming distance /Length of string

C. Hamming distance * Length of string

D. Hamming distance + Length of string

Q11. Which of the following correctly explains the Levenshtein distance?

A. Number of positions with the same symbol in both strings.

B**. Minimum number of insertions, deletions, and replacements needed for transforming string A into a string B.**

C. Minimum number of symbols to be removed in both strings until resulting substrings are identical

D. Sum of absolute differences between N-gram vectors of both strings.

Q12. What is the Hamming distance between the binary vectors a =1101011111 and b = 0111011000?

A. 3 B**. 5** C. 6 D. 2

Q13. Choose the correct hamming distance for the following two strings: "Caroline" and "Cathrine".

A. 5 B. 2 C. **3** D. 1S

Q14. Which of the following is/are the application(s) of hamming distance?

i. To calculate the genetic distance between genotypes of the virus

ii. For detecting and repairing wrong bits.

iii. To calculate the number of characters that need to be replaced to transform one string into another

iv. In identifying user intent from a given statement

A. Only ii and v B. Only i, ii, and iii c. Only ii D. Only ii and v

Q.16 What is the hamming distance of the following pairs of strings: "rocket, rockstar", and unbuttoned, unbuttered"?

A. 1,2

B. 3,2

C. 2,1

D. Cannot be calculated, 2

Q.17 Calculate the Hamming distance and Levenshtein distance for the following two strings: "man and "woman".

A. Hamming distance =2, Levensthein distance = 2

B. Hamming distance =0, Levensthein distance = 0

C. Hamming distance =2, Levensthein distance = 0

D. Hamming distance =Cannot be calculated, Levenshtein distance = 2

Q.18 The Levenshtein distance is an edit distance-based algorithm as it computes the number of edits required to transform one string to another. Choose all the options that the edits count as an operation.

A. Insertion of a character

B. Deletion of a character

C. Changing the case of a character

D. Substitution of a character

Q19. Compute the Levenshtein distance for the two strings: "transform" and "transmit".

A. 5

B. 4

C. 1

D. 3

Q20. Hamming distance is inversely proportional to the similarity of two strings. True or False?

True

False

Q21. What is the Hamming distance between the binary vectors a =10111011001 and b= 01110011011?

A. 2

B. 7

C. 3

D. 4

Which of the following is not a part of feature engineering?

A. Creating a new feature to determine weekend or not from the dates fea

B. One-hot encoding gender feature.

C. Creating new feature for kph speed from mph speed.

D. All of the above

Choose all the text representation techniques.

Multiple answers are possible.

A. TF-IDF

B. Target encoding

C. Bag of words

D. All of the above

Creating month feature from date feature is feature engineering technique.

A. True

B. False

Consider the following corpus.

'I', 'love', 'city', 'in', 'data', 'engineering', 'nlp', 'and'

What is the bag of words representation of the following sentence?

"I love data engineering"

A. [1, 1, 1, 1, 1, 1, 1, 1]

B. [1, 1, 0, 0, 1, 1, 0, 1]

C. [1, 1, 0, 0, 0, 0, 1, 0]

D. [1, 1, 0, 0, 1, 1, 0, 0]

Skip-gram Word2Vec model tries to predict the neighbours of a word.

A. True

B. False

Which of the following is not a valid bigram phrase selected from the sentence below?

"I love NLP and Data Science"

A. I love

B. Data Science

C. I Science

D. Love NLP

Consider the following documents.

Document 1 : I love apples

Document 2 : I hate oranges but I love kiwis

What is the IDF value of "love" in Document 2?

A. 1

B. 0

C. 0.1

D. 10

Which of the following sentences is not true regarding language modeling?

A. Language modeling is the task of assigning a probability to a sentence or a phrase.

B. Language modeling can be represented as P(S) = P(w1, w2, w3, ..., wn), where w1, w2, w3, ..., wn are occurring together.

C. Language modeling is not used to compute the probability of upcoming words.

D. The probability of a sentence in language modeling means how likely it is to occur in natural

How will you compute the probability of upcoming words in language modeling?

A. P(Wn-1 | W1, W2, ..., Wn)

B. P(Wn | W1, W2, ..., Wn-1)

C. P(Wn | W1, W2, ..., Wn)

D. None of the above

Select all the applications of language modeling in natural language processing.

Multiple answers are possible.

A. Text classification

B. Search engines

C. Auto-correct feature

D. Machine translation

Which of the following options is the correct way to compute conditional probability?

A. $P(A \mid B) = P(A, B) / P(B)$

B. $P(A \mid B) = P(A, B) * P(B)$

C. $P(A \mid B) = P(A, B) / P(A)$

D. None of the above

What is the probability of the sentence using conditional probability and chain rule: "eat this apple"?

A. P(eat, this, apple) = P(apple) + P(this | apple) + P(eat | apple, this)

B. P(eat, this, apple) = P(eat) + P(this | eat) + P(apple | eat, this)

C. P(eat, this, apple) = P(eat) * P(this | eat) * P(apple | eat, this)

D. P(eat, this, apple) = P(apple) * P(this | apple) * P(eat | apple, this)

State true or false: "In Markov assumption (in language modeling), it says that instead of looking at the entire context, let's just look at the past few words which will be my most relevant context for that particular word."

A. True

B. False

Select the correct bigram for the following sentence: "how are you doing".

A. ["how", "are", "you", "doing"]

B. ["how are you", "are you doing"]

C. ["how are", "are you", "you doing"]

D. None of the above

Which of the following statements is true with respect to bigrams?

A. The number of bigrams in a sentence will be one more than the number of tokens in that sentence.

B. The number of bigrams in a sentence will be one less than the number of tokens in that sentence.

C. The number of tokens in a sentence will be one less than the number of bigrams in that sentence.

D. None of the above

How many trigrams are in the following sentence: "how are you doing"?

A. 1

B. 2

C. 3

D. 4

Perplexity is one of the commonly used evaluation metrics for the evaluation of language models. Under which category does perplexity fall?

A. Extrinsic evaluation

B. Intrinsic evaluation

C. None of the above

Using Conditional Probability and Chain Rule, what is the probability of the phrase: "the day looks windy"?

A. P(the) * P(day|the) * P(looks|the, day) * P(windy|the, day, looks)

B. P(the) * P(the|day) * P(the, day|looks) * P(the, day, looks|windy)

C. P(the|day, looks, windy) * P(day|looks, windy) * P(looks|windy) * P(windy)

D. P(the, day, looks|windy) * P(day, looks|windy) * P(looks|windy) * P(windy)

What is the correct order of steps involved in Text Classification?

A. Feature Engineering - Classify the text

B. Text Preprocessing - Feature Selection - Feature Engineering - Classify the text

C. Text Preprocessing - Feature Engineering - Classify the tex

D. None of the above

State True or False: "Hand-coded model of text classification uses Machine Learning models to classify text"

A. True

B. False

Which of the following is(are) not true about Hand-coded rules?

i. Rules are based on domain knowledge

ii. Rules are always going to be non-exhaustive

iii. Rules can be generalized

iv. Hand-coded Model is difficult to interpret

A. i, ii

B. iii, iv

C. ii, iii, iv

State True or False: "In Hand-code model, the lesser rules we have, the higher the accuracy of the model"

A. True

B. False

How will a Hand-coded model classify the following review based on positive and negative words only?

Review: “The food is awesome and the service is not bad. People not visiting this place are missing out on good food”

Note: The non-exhaustive list of positive and negative words only contains phrases with single words and are domain-specific

A. Positive review

B. Negative review

State True or False: “The reason behind Naive in Naive Bayes is that it assumes all the features to be independent of each other”

A. True

B. False

Which of the following is(are) true about the Machine Learning way of text classification?

i. No manual intervention is required to make rules for classification

ii. This method requires the input set of data to be labeled for training

iii. Using documents that the model has seen previously as the test dataset is a good way to evaluate the model.

iv. The data set used for evaluation is much bigger than the training data set.

A. i, ii

B. i, ii, iii

C. ii, iv

Q. Rapid changes in image intensity are caused following phenomena.
A. Surface Normal Discontinuity
B. Depth Discontinuity
C. Illumination Discontinuity
D. All of the Above

Q. Which one is not a valid type of edge?
A. Step Edge
B. Roof Edge
C. Line Edge
D. Circle Edge

Q. Which of the following is not a job of an Edge Operator/detector?
A. Detecting Edge Position
B. Detecting Edge Magnitude
C. Detecting Edge color
d. Detecting Edge Orientation

Q. The general criteria for edge detection include:

A. Low error rate
B. Good localization
C. Minimal response
D. All of the above

Q. Which one is correct for the Laplacian method?
A. Provides direction of the edge.
B. Detection based on Zero-Crossing.
C. Non-linear operation.
D. Detection using Maxima Thresholding

Q. Which of the following is not belong to the Canny edge detection algorithm?
A. Apply the Laplacien filter to smooth the image in order to remove the noise
B. Find the intensity gradients of the image.
C. Apply non-maximum suppression to get rid of spurious response to edge detection.
D. Apply a double threshold to determine potential edges.

Which of the following definition defines Boosting?

A. Boosting is a classification scheme that works by combining weak learners into a more accurate ensemble classifier

B. During each boosting round, we select a strong learner that does well on hard examples for the previous learners.

C. During each boosting round, we select a weak learner that does well on examples that were easy for the previous strong learners

D. Boosting is a classification scheme that combines strong learners into a more accurate ensemble classifier.

Q. Which of the following is not a correct reason to use Bag-of-word/Bag-of Features/bag-of-shape in computer vision?

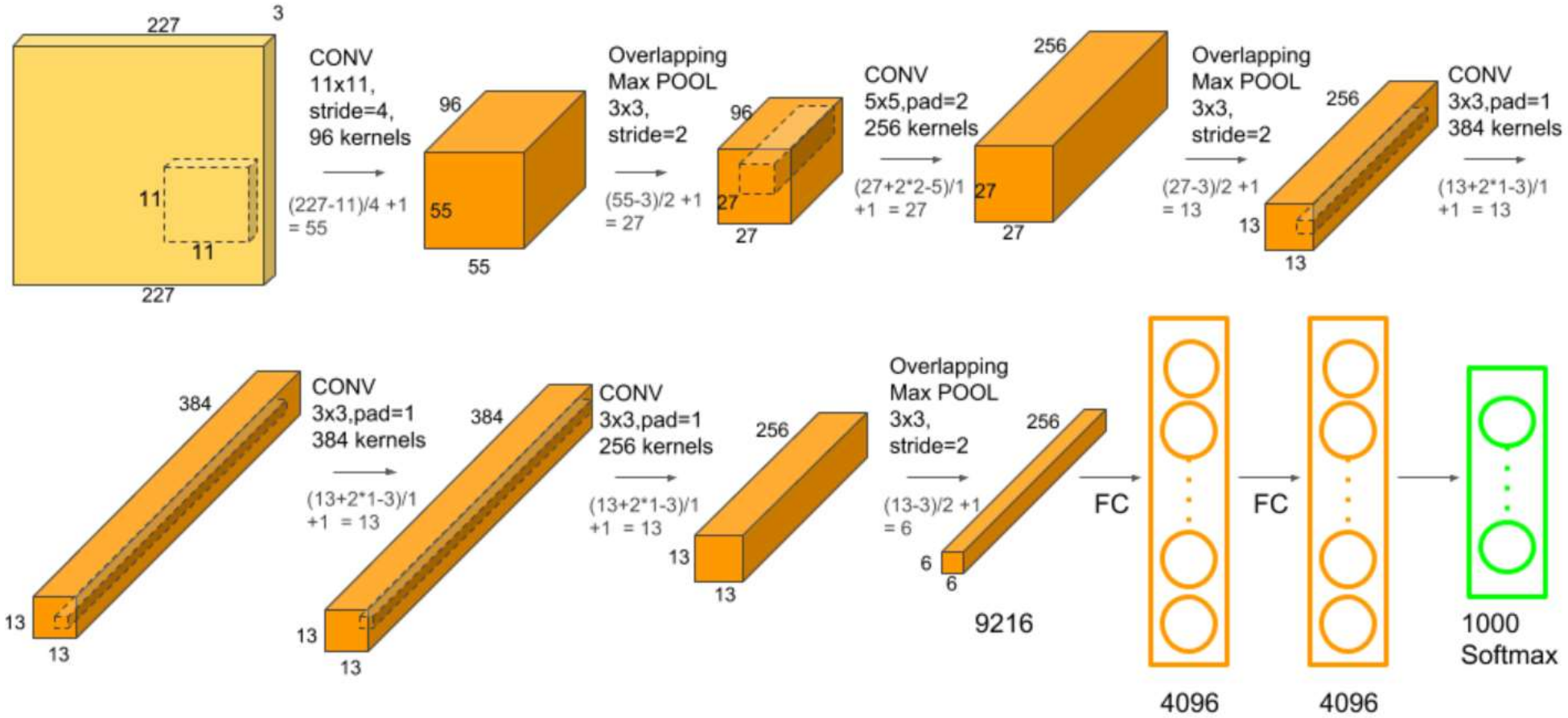
A. Learn “visual vocabulary”

B. Quantize features using a visual vocabulary

C. Represent images by frequencies of “visual words”

D. Apply ML models to Image

Q. Below architecture represents:



A. AlexNet
B. ResNet
C. ImageNet
D. GoogleNet

If your computer vision model identifies a painting as an object class, and “Mona Lisa” by Leonardo Da Vinci is an instance of that painting then it is known as.

A. CNN model
B. Instance Level Recognition
C. Object Level Recognition
D. RNN Model

YOLO algorithm is used to detect-------

A. Objects in real-time.
B. Offline Image processing
C. Edge Detection
D. Color Detection

-------- is only an encoder, while the original transformer is composed of an encoder and decoder

A. Transformer
B. GRU
C. LSTM
D. BERT

When there are more layers in the network, the value of the product of the derivative decreases until at some point the partial derivative of the loss function approaches a value close to zero. This is known as ----------.

A. Vanishing gradient problem.
B. Gradient Descent
C. LSTM
D. Attention Layer

Which of the following option is used You Only Look Once approach?

A. Fast RNN
B. Fatter RNN
C. YOLO
D. Sliding Window