Recommender Systems

Assignment- Week 5

TYPE OF QUESTION: MCQ/MSQ

Number of questions: 20 **Total marks:**

QUESTION 1:

One of the advantage of the content-based recommendation system is

a) It can recommend items to new users without any difficulty
b) It can recommend new items to existing users without any difficulty
c) It can avoid both user and item cold start problems
d) It does not require the item features for generating recommendations.

Correct Answer: b

Explanation: Refer to Week 5 lecture 1 slide 5

QUESTION 2:

Content based recommender systems use prediction models for recommending items to the users. In these models the independent variables are _____ and the dependent variable is _____.

a) Item features, Preference rating by the particular user
b) Preference rating by the particular user, Item features
c) All Item feature, the target item features
d) All preference rating by the users, the preference rating for a target item

Correct Answer: a

Explanation: Refer to Week 5 lecture 1 slide 9

QUESTION 3:

Process of finding the features from raw data, selecting the best set of features that makes the prediction accuracy better, and transforming the original set of features to a lower dimensional space are termed as _____, _____ and _____ respectively.

a) Feature Selection, Dimensionality reduction, Feature extraction
b) Feature Selection, Feature extraction, Dimensionality reduction
c) Feature extraction, Dimensionality reduction, Feature Selection
d) Feature extraction, Feature Selection, Dimensionality reduction

Correct Answer: d

Explanation: Refer to Week 5 lecture 2 slide 4

QUESTION 4:

During feature extraction the raw data is transformed in to _____form for further processing.

a) Text
b) Numeric
c) Dimension
d) Feature

Correct Answer: b

Explanation: Refer to Week 5 lecture 2 slide 5

QUESTION 5:

During text feature extraction, the process of splitting a text documents into words is called _____.

a) Stemming
b) Lemmatization
c) Feature extraction
d) Tokenization

Correct Answer: d

Explanation: Refer to Week 5 lecture 3 slide 4

QUESTION 6:

WordNet is an example of a database of _____

a) Root words to be used during lemmatization
b) Stop words
c) Named entities
d) Lexical relations

Correct Answer: d

Explanation: Refer to Week 5 lecture 3 slide 13

QUESTION 7:

Which of the following represents the entropy associated with a random variable X

a) $\log_2 p(x)$
b) $-\log_2 p(x)$
c) $-\Sigma_x p(x) \log_2 p(x)$
d) $\Sigma_x p(x) \log_2 p(x)$

Correct Answer: c

Explanation: Refer to Week 5 lecture 4 slide 5

QUESTION 8:

Which of the following is the correct formula to calculate the mutual information I(X,Y) between two random variables X and Y. Please note that H(.) represents entropy.

a) H(X) - H(X|Y)
b) H(Y) – H(Y|X)
c) H(X) +H(Y) – H(X,Y)
d) all of the above

Correct Answer: d

Explanation: Refer to Week 5 lecture 4 slide 7, 11

QUESTION 9:

A customers binary rating data suggest, out of 10 products he has rated 5 as positive and 5 as negative. What is the Gini index value associated with the rating variable?

a) 0
b) 1/2
c) 3/4
d) 1

Correct Answer: b

Explanation: Refer to Week 5 lecture 5 slide 3-4

QUESTION 10:

In a wrapper based feature selection approach which of the following is the right sequence of activities

a) Initial subset selection, Trying the subset with a learner, Measuring the difference, Modifying the feature subset, Trying again
b) Initial subset selection, Modifying the feature subset, Trying the subset with a learner, Measuring the difference, Trying again
c) Initial subset selection, Measuring the difference, Trying the subset with a learner, Modifying the feature subset, Trying again
d) Initial subset selection, Trying the subset with a learner, Modifying the feature subset, Measuring the difference, Trying again

Correct Answer: d

Explanation: Refer to Week 5 lecture 5 slide 7

**************END*************