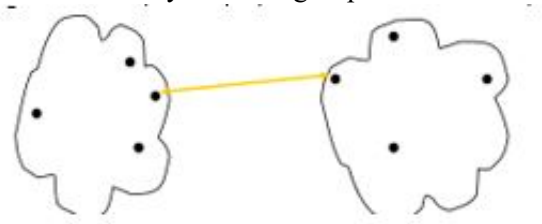
قائمة الاسئلة

تنقيب البيانات -المستوى الرابع -علوم حاسوب - الكل - كلية الحاسوب وتكنولوجيا المعلومات - درجة الامتحان (40)

أ.هبه المروعي

1) The k-means clustering algorithm that we studied will automatically find the best value of k as part of its normal operation
   1) - TRUE.
   2) + FALSE.
2) In association rule the generation of the frequent itermsets is the computational intensive step
   1) + TRUE.
   2) - FALSE.
3) Any subset of a frequent set is a frequent set
   1) + TRUE.
   2) - FALSE.
4) The ---------- step eliminates the extensions of (k-1)-itemsets which are not found to be frequent, from being considered for counting support
   1) - a)Candidate generation
   2) + b)Pruning
   3) - c)Partitioning
   4) - c)A&B
5) One easy way to reduce SSE is to reduce K (number of clusters)
   1) - TRUE.
   2) + FALSE.
6) Outliers are always the cause of noise
   1) - TRUE.
   2) + FALSE.
7) In the figure below, there are two groups. They are connected to a line that represents the distance used to determine similarity between groups What measure of similarity between groups does this line represent?



   1) - Max
   2) + Min
   3) - Distance between centroids
   4) - Group Average
8) Which one of the clustering technique needs the merging approach
   1) - a)Partitioned
   2) - b)Naïve Bayes
   3) + c)Hierarchical
   4) - Both A and C
9) duplicate is the main reason for inconsistency problem
   1) + TRUE.
   2) - FALSE.
10) In sampling with replacement, the same object can be picked up more than once
   1) + TRUE.

2) - FALSE.

11) The value of IDF(w) is minimized when the word is common to all documents
1) + TRUE.
2) - FALSE.

12) Inverse document frequency IDF(w) is a natural measure of the uniqueness of the word w
1) + TRUE.
2) - FALSE.

13) When the value of IDF(w) = 0 that means the word is unique to the document
1) - TRUE.
2) + FALSE.

14) $(C_1, C_2) = |C_1 \cap C_2| / |C_1 \cup C_2|$.

is ------formula of
1) + Jaccard Similarity
2) - Intersection Similarity
3) - Cosine Similarity
4) - None of the above

15) if the value of cos(X,Y) =1 that means the two vectors are…..
1) + aligned
2) - Orthogonal
3) - intersection
4) - None of the above

16) distance is numerical measure of how …….two data objects are
1) - a like
2) + different

17) A good clustering ……Intra-cluster distances
1) + minimized
2) - maximized

18) Exclusive clustering is…….
1) - a)Partitional Clustering
2) - b)None fuzzy clustering
3) - c)fuzzy clustering
4) + a&b

19) K-means Clustering is ……
1) - a)Partitional Clustering
2) - b)None fuzzy clustering
3) - c)fuzzy clustering
4) + a&b

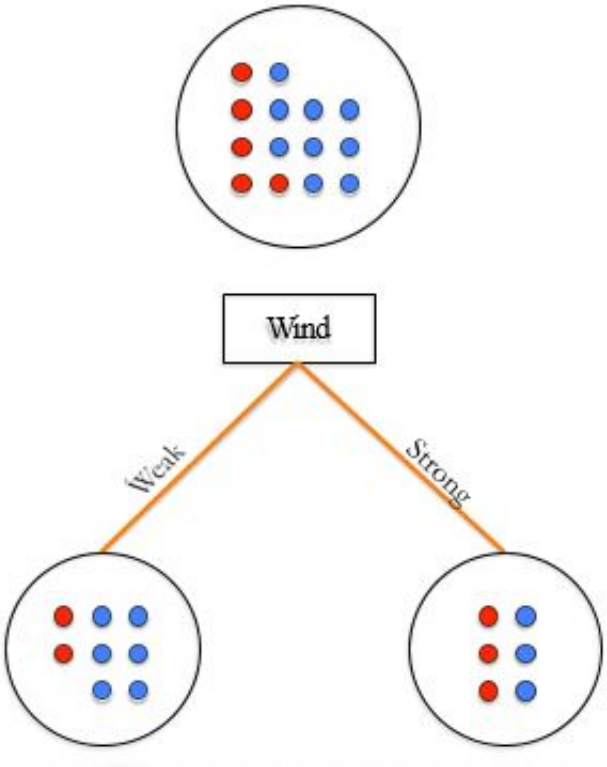
20) A good clustering with …… K can have a ……. SSE
1) + smaller,lower
2) - smaller,higher
3) - larger,lower
4) - larger,higher

21) The best way to determine initial centroids for k-means algorithms is
1) - a)use hierarchical clustering
2) - b)run algorithm more than times
3) - c)use fuzzy clustering
4) + a&c

22) The entropy of the fallowing attribute is ….

1) - 1
2) - 0
3) - 0.5
4) + 0.95

23) What is the primary objective of the k-means clustering algorithm?



1) - A) Minimize the distance between points in the same cluster
2) - B) Maximize the distance between clusters
3) + C) Both A and B
4) - D) None of the above

24) Given a threshold of 0.2 and a learning rate of 9.1, along with the following information, what will be the value of the new weight q24.png?

Epoch	Inputs		Desired output	Initial weights		Actual output
	x_1	x_2	Y_d	w_1	w_2	Y
	1	0	0	0.2	0.1	1

1) - 0.3,-0,1
2) - 0.3,0.0
3) - .0.2,0.1
4) + 0.1,0.1

25) The value of sign activation function is -----if x>=0
1) + a)1
2) - b)0
3) - c)-1
4) - a&b

26) In k-means, what does SSE (Sum of Squared Errors) represent?
1) - A) The distance between centroids
2) + B) The total distance of each point from its assigned centroid
3) - C) The number of clusters
4) - D) The time complexity of the algorithm

27) What is the main purpose of a decision tree algorithm?
1) - A) To reduce dimensionality
2) + B) To classify or predict outcomes based on input features
3) - C) To cluster similar data points
4) - D) To visualize data distributions

28) What happens during the "pruning" step of the Apriori algorithm?
1) + A) Non-frequent itemsets are removed
2) - B) Frequent itemsets are expanded
3) - C) Duplicate transactions are eliminated
4) - D) Rules with low confidence are discarded

29) What is the first step in the Apriori algorithm?
1) - A) Generate rules from frequent itemsets
2) + B) Scan the database to count item frequencies
3) - C) Calculate lift for association rules
4) - D) Create candidate itemsets

30) In the context of the Apriori algorithm, what does a "frequent itemset" signify?
1) - A) An itemset that appears in less than the minimum support threshold
2) + B) An itemset that appears in at least the minimum support threshold
3) - C) Any combination of items
4) - D) A single item that is frequently purchased

31) What is the purpose of data preprocessing in data mining?
1) - A) To create visualizations
2) + B) To prepare and clean data for analysis
3) - C) To store data in a database
4) - D) To perform clustering

32) In data mining, what does "association rule mining" focus on?
1) - A) Predicting future values
2) + B) Finding relationships between variables in large datasets
3) - C) Segmenting data into clusters
4) - D) Reducing dimensionality

33) What does the “k” in k-means stand for?
1) - A) The number of dimensions
2) + B) The number of clusters
3) - C) The number of iterations
4) - D) The number of data points

34) In data mining, data visualization is not important.
1) - TRUE.
2) + FALSE.

35) Neural networks are a type of unsupervised learning algorithm.

1) - TRUE.
2) + FALSE.

36) Data mining can be applied to both numerical and categorical data.
1) + TRUE.
2) - FALSE.