

APJ ABDULKALAM TECHNOLOGICAL UNIVERSITY
08 PALAKKAD CLUSTER

Q. P. Code : 08CSE19-6022-1

(Pages: 3)

Name:

Reg. No:.....

SECOND SEMESTER M.TECH. DEGREE EXAMINATION MAY 2019

Branch: Computer Science and Engineering Specialization: Computer Science and Engineering

08CS 6022 INFORMATION RETRIEVAL

(Common to CSE)

Time:3 hours

Max. marks: 60

Answer all six questions.

Modules 1 to 6: Part 'a' of each question is compulsory and answer either part 'b' or part 'c' of each question.

| Q.no. | Module 1 | Marks |
|----------------------|--|--------------|
| 1.a | Compare the process of data retrieval and information retrieval. | 3 |
| Answer b or c | | |
| b | Write note on Recall and Precision for retrieval performance evaluation. Explain with a simple example why Recall or Precision alone is not enough to quantify the performance of IR System. Explain any two alternative measures which combine recall and precision to get a better performance evaluation. | 6 |
| c | Describe the Retrieval process with simple, generic software architecture. | 6 |
| Q.no | Module 2 | Marks |
| 2.a | Explain how Euclidean distance function and cosine similarity can be used as similarity measures. Will a document pair found 'similar' with cosine similarity measure be similar with distance function similarity measure? Justify your answer with a suitable example. | 3 |
| Answer b or c | | |
| b | Write note on Boolean and Vector classical models for information retrieval. What are the advantages of vector model over Boolean model. | 6 |
| C | Explain the classic probabilistic model. What are the advantages and disadvantages of this model? | 6 |
| Q.no. | Module 3 | Marks |
| 3.a | Discuss Page Ranking. | 3 |

Answer b or c

- b Write note on the User Relevance Feedback strategy for query reformulation. **6**
 Explain the process of Query Expansion and Term Reweighting for vector model.
- c Write note on the process of query expansion through Local Clustering. Explain **6**
 three cluster building strategies for local clustering in detail (association clusters, metric clusters, and scalar clusters).

| Q.no. | Module 4 | Marks |
|-------|---|-------|
| 4.a | Write note on two type of multimedia similarity queries <i>Whole Match</i> and <i>Sub-pattern Match</i> queries with example. | 3 |

Answer b or c

- b Why we need spatial access methods instead of sequential scanning to access **6**
 multimedia objects? Explain GEMINI algorithm (Explain each step).
- c What do you mean by *curse of dimensionality*? Discuss the difference between **6**
 feature selection and feature extraction with example. How do these two process contribute to dimensionality reduction?

| Q.no. | Module 5 | Marks |
|-------|--|-------|
| 5.a | Discuss steps involved in decision tree classification process with a simple example .List its advantages and disadvantages. | 4 |

Answer b or c

- b Discuss k-NN and Naive Bayes classification. In the figure given below two **8**
 types of two-dimensional data distributions are given. One is linearly separable and other is linearly non-separable. You are asked to select a classifier from kNN and Naive Bayes for each of data distribution. Justify your selection.

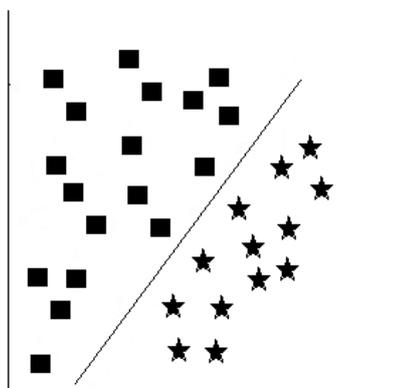


fig:1

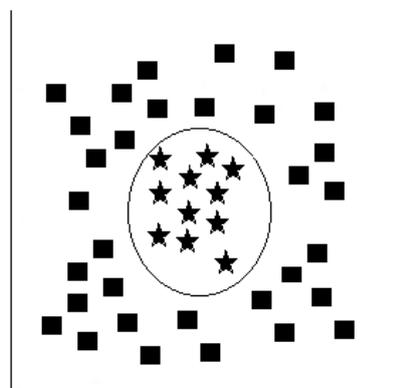


fig:2

- c Explain Naive Bayes classifier. With Naive Bays classifier and data given in table below compute $P(Yes|x')$, $P(No|x')$ for the unseen data $x'=(Outlook=Sunny, Temperature=Cool, Humidity=High, Wind=Strong)$. How will you interpret the result? **8**

| Day | Outlook | Temperature | Humidity | Wind | PlayTennis |
|-----|----------|-------------|----------|--------|------------|
| D1 | Sunny | Hot | High | Weak | No |
| D2 | Sunny | Hot | High | Strong | No |
| D3 | Overcast | Hot | High | Weak | Yes |
| D4 | Rain | Mild | High | Weak | Yes |
| D5 | Rain | Cool | Normal | Weak | Yes |
| D6 | Rain | Cool | Normal | Strong | No |
| D7 | Overcast | Cool | Normal | Strong | Yes |
| D8 | Sunny | Mild | High | Weak | No |
| D9 | Sunny | Cool | Normal | Weak | Yes |
| D10 | Rain | Mild | Normal | Weak | Yes |
| D11 | Sunny | Mild | Normal | Strong | Yes |
| D12 | Overcast | Mild | High | Strong | Yes |
| D13 | Overcast | Hot | Normal | Weak | Yes |
| D14 | Rain | Mild | High | Strong | No |

PlayTennis: training examples

| Q.no. | Module 6 | Marks |
|-------|---|-------|
| 6.a | Write note on Recommender Systems and its applications. | 4 |

Answer b or c

- b Explain K-Means clustering algorithm. Do think k-means clustering is a special case of hard EM(Expectation Maximization).Justify your answer. List advantages and disadvantages of k-means clustering. Analyse the time complexity of K-Means clustering. **8**
- c Explain process of Hierarchical agglomerative clustering with a suitable example. How it is different from divisive approach? Analyse its time complexity. **8**