

# M.Sc. CS/MCA Semester - III Examination 2025

## CS-209 Machine Learning (Lab)

(Answer any 2 questions)

**Time: 3 hrs**

**F.M: 30**

1. You are given the MNIST handwritten digits dataset (70,000 grayscale images of size 28×28).

Write Python code to:

- a) Load the MNIST dataset and flatten the images into feature vectors. (3)
- b) Apply PCA to reduce the dimensionality to 50 principal components. (4)
- c) Train a Logistic Regression classifier on the reduced feature space. (4)
- d) Evaluate the classifier on the test set and print the accuracy. (4)

Python script to load dataset:

```
from sklearn.datasets import fetch_openml  
  
mnist = fetch_openml("mnist_784", version=1, as_frame=False)
```

2. Using the Iris benchmark dataset (UCI repository), write Python code to:

- a) Load the dataset and split it into train and test sets (80/20). (2)
- b) Standardize the features using StandardScaler. (2)
- c) Train a K-Nearest Neighbors classifier ( $k = 5$ ) on the standardized data. (3)
- d) Train a Gaussian Naive Bayes classifier on the same standardized data. (3)
- e) Compare the accuracy of the two models on the test set and print which classifier performs better. (5)

Python script to load dataset:

```
from sklearn.datasets import load_iris
```

3. Using the Wine dataset (UCI repository), write Python code to:

- a) Load and standardize the dataset. (2)
- b) Apply PCA to reduce the dataset to 2 principal components for visualization. (3)
- c) Perform Agglomerative Clustering with 3 clusters and plot the resulting clusters in PCA space. (5)
- d) Perform K-Means clustering with  $k = 3$  and plot its clusters on the same PCA-transformed space. (5)

Python script to load dataset:

```
from sklearn.datasets import load_wine
```

## M.Sc. Semester III Examination 2025-26

Computer Science

## CS-207: Compiler Design

Time : Three hours

Max. Marks : 70

(Write your Roll No. at the top immediately on the receipt of this question paper)

**Note:** Attempt five questions from the following including Question No. 1, which is compulsory.

- |   | Marks  |
|---|--------|
| 1/ (a) Explain the process of tokenization in lexical analysis.   | [3.5×4 |
| (b) What is the difference between LL(1) and LR(1) parsing techniques?  | =14]   |
| (c) Describe the symbol table data structure used in compiler design and its significance during compilation. |        |
| (d) What is the purpose of generating intermediate code in a compiler?  |        |
|   |        |
| 2/ (a) Identify the Tokens, Pattern and Lexemes in the following code segment                                 | [7]    |
| <pre> int main )( /*Example*/ fro(i=1: i&lt;=20; i++) x=b*c++; printf("%d%d",i, x) } </pre>                   |        |
| Does lexical analyzer generate any error/recovery in this code during compilation?                            |        |
| (b) What are the various phases of Compiler Design? Discuss each in brief.                                    | [7]    |
|   |        |
| 3 (a) Consider the following grammar:   | [6]    |
| $A \rightarrow A+A \mid A*A \mid id$  |        |
| i. If the grammar is ambiguous then convert it to unambiguous grammar.  |        |
| ii. Remove left recursion from the grammar obtained at step i   |        |
| iii. Give a parse tree for string '7+4*2' using grammar obtained at step ii.                                  |        |
|   |        |
| (b) Construct SLR(1) parsing table for the following grammar:   | [8]    |
| $S \rightarrow OS0 \mid IS1 \mid 0$   |        |
| And parse an input string 10001.  |        |
|   |        |
| 4/ (a) Discuss the differences between top-down and bottom-up parsing techniques                              | [7]    |
| in the context of syntax analysis. Provide examples of parsers for each approach.                             |        |

P.T.O.



(2)

- (b) Construct a predictive parsing table for the following grammar: [7]

$S \rightarrow A$   
 $A \rightarrow aB \mid A$   
 $B \rightarrow bBC \mid f$   
 $C \rightarrow g$

- 5 (a) What is Handle? Consider the following grammar, and show the handle of each right sentential form for the string  $(a, a)$  [6]

$S \rightarrow (A) \mid \lambda$   
 $A \rightarrow A, S \mid S$

- (b) What is an operator precedence parser and operator grammar? Compute LEADING and TRAILING in the following grammar [8]

$P \rightarrow P + F \mid F$   
 $F \rightarrow F * G \mid G$   
 $G \rightarrow (P) \mid x$

- 6 (a) What is Syntax Directed Definition (SDD)? Consider the grammar with productions [4]

$S \rightarrow TL$   
 $T \rightarrow int$   
 $L \rightarrow L, id$   
 $L \rightarrow id$

Give semantic rules and annotated parse tree for the declaration

$int\ w, x, y$

- (b) Generate Three address code for the following code segment: [4]

if  $(x > y)$   
     $z = x - y;$   
else  
     $z = y - x;$

- (c) Translate the arithmetic expression  $(a+b)*(c-d)+(a+b)-c*d$  into

- A syntax tree
- DAG
- Quadruples

[6]

- 7 (a) Explain the concept of code optimization and provide examples of common code optimization techniques used by compilers. [7]

- (b) What are basic blocks and flow graphs? Why is it needed? Write an algorithm to partition three address code into basic blocks. [7]

— 0 —

**M.Sc./M.C.A. Semester III Examination 2025-26****CS-209: Machine Learning****Time : Three hours****Max. Marks : 70****(Write your Roll No. at the top immediately on the receipt of this question paper)**

Note: Attempt any five out of seven given questions. Marks of each question are written on the right side.

- Q1 a) What is the difference between underfitting and overfitting the training data? 7  
b) Explain one hot encoding. What is dummy variable trap? Explain it with an example. 7
- Q2 a) Explain data pre-processing framework, in general, for any classification problem. 7  
b) What is the confusion matrix? Explain any two metrics to evaluate machine learning models. 7
- Q3 a) What is the variance inflation factor (VIF)? Explain its role in regression models. 7  
b) Explain the backward elimination method of model selection in cases of a large number of variables/features. 7
- Q4 a) What is the role of regularization? Explain different types of regularization. 7  
b) What do you understand by confounding and mediator variables in context of multiple linear regression. Explain by examples. How do we control these variables? 7
- Q5 a) Explain how information gain is used to select attributes for decision tree splitting by an example. 7  
b) Explain Chi-square method for decision tree splitting with a suitable example. 7
- Q6 a) Explain fuzzy c-means clustering using an example. 7  
b) What are the drawbacks of k-means algorithm? How we address it in K-means++ algorithm? 7
- Q7 a) What do you understand by multi-dimensional data? Explain any one technique for dimensionality reduction. 7  
b) Define Bayes minimum error and Bayes minimum risk classifier. In which case both will be the same? 7



**M.Sc./M.C.A. Semester III Examination 2025-26****CS-353: Internet of Things****Time : Three hours****Max. Marks : 70**

---

**(Write your Roll No. at the top immediately on the receipt of this question paper)**

---

**Note:** The question Paper contains 06 questions out of which you are required to answer any 05 questions. The question paper is of 70 marks with each question carrying 14 marks.

**Note:** Attempt any five questions.

**Time: Three Hours****MM-70**

1. (a) Discuss the genesis of IoT. Define IoT according to the IEEE and other benchmarks with illustrations. Discuss the key points also.  
(b) Discuss different types of actuators and sensors with suitable examples.
2. (a) Discuss the major scalability issues faced when deploying IoT systems on a large scale. Explain the difference between horizontal and vertical scalability in IoT.  
(b) Define IoT security and explain why it is critical in IoT application development. List any four common security threats in IoT applications.
3. (a) Discuss the five-layer IoT architecture and its advantages over the traditional model.  
(b) Design and explain a layered IoT architecture for a smart city application. Discuss the data flow and interaction between layers.
4. (a) Define IoT gateways. Discuss the major functions of an IoT Gateway such as protocol translation, data filtering, and device management.  
(b) Design an IoT-based wearable system for patient monitoring. Explain its architecture, sensors, communication flow, and data analytics.

**PTO**

(2)

5. (a) Explain communication criteria range and frequency bands of IoT networks.

(b) Explain Edge computing and Edge Analytics and its environment.

6. Write notes on any two of the followings.

a. IoT Stack

b. Machine Learning for IoT

c. IoT data Management

d. IoT in Public safety

— 0 —



REFNO0.....

Roll No. ....

**M.Sc./M.C.A. Semester III Examination 2025-26**

**Computer Science**

**CS-312: Information Retrieval**

**Time : Three hours**

**Max. Marks : 70**

---

(Write your Roll No. at the top immediately on the receipt of this question paper)

**Note:** Attempt total Five question including compulsory question 1. The figures in right margin indicate the marks.

1. (a) A Boolean index of a corpus of documents is given represented as List. Explain how to search for the query :

'Information' AND 'Retrieval' NOT 'System'

- (b) Two inverted indices (INDEX1 and INDEX2) of two different corpuses are given as two matrices of weights. INDEX1 has 100 Words and 50 documents, while INDEX2 has 75 Words and 125 documents. There are 25 common Words in both indices, but no common documents in both corpuses. Create a new INDEX3 by appropriately merging the two given indices. Mention all the dimensions of the indices. Also mention clearly how to calculate the weights of INDEX3.

- (c) Write and explain the 'Probability Ranking Principle' ( PRP ). Briefly explain its use in Probabilistic IR Systems.

2. (a) Suppose a User has not much idea about a topic, e.g. 'Agricultural Economics'. The user decides to iteratively improve the query by adding few words and removing few words of the query. Write an appropriate algorithm for this task. Explain each step by giving an example.

- (b) What do you mean by 'Focus - Shift '. Explain through an example in the above task.

3. If a corpus has documents from diverse domains, such as, Sports, Entertainment, Computer Science, AI, etc., the distribution of words in the corpus may follow a Gaussian Mixture Model. To design an IR system for such a corpus we can cluster the Corpus first and apply the queries on few clusters. Write an algorithm for this task and explain the document search process with an example query.

4. (a) What is the source of uncertainties in an IR System and how can one resolve some of them in the Probabilistic IR Model. 7
- (b) How do you define 'Retrieval Status Value'(RSV). Give the mathematical formulae for it. Explain every term in the mathematical formulae and describe briefly how to get these values practically to design an IR System. 7
5. (a) Write steps to automatically find out the stop words from the given corpus in any one or multiple languages. 7
- (b) Compare the Vector Space Model and the Probabilistic Model of an IR System. 7
6. (a) Explain with examples a set of 7 important computable Image Features for Image Retrieval. 7
- (b) Compare the Content and Context Based Image Retrieval with examples. 7
7. Write short notes on any two of the following : 7x2 =14
- i. F-measure and E-measure.
  - ii. Latent Semantic Indexing.
  - iii. Use of skip pointers for searching in boolean IR Models.