

2025

M.Sc. 1st Semester Examination

COMPUTER SCIENCE

Paper : COSE405A1 / B1 / C1

Full Marks : 35

Time : One Hour Thirty Minutes

*The figures in the margin indicate full marks.
Candidates are required to give their answers
in their own words as far as practicable.*

Paper : COSE405A1

(Data Science)

Group - A

Answer any *five* questions : $2 \times 5 = 10$

1. Give examples of one partition based clustering and one density based clustering algorithm.
2. Why Euclidean distance is considered as a similarity index while clustering?
3. Why is regression technique considered a supervised machine learning algorithm?
4. List any four applications/use cases of Data Science.
5. Name any four data types or data structures commonly used in Data Science.

P.T.O.

(2)

6. Define probability distribution. Name any two probability distributions. 1+1
7. A manufacturing process claims a mean output of 50 units, explain how hypothesis testing can be used to verify this claim.
8. Differentiate between Supervised and Unsupervised learning. Give one example each. 1+1

Group - B

Answer any *three* questions : $4 \times 3 = 12$

9. Explain the different kinds of Skewness a data distribution can possess.
10. Briefly explain the basic steps of data preprocessing.
11. What is an outlier? Explain the boxplot with the quartiles. 1+3
12. (a) How does a *numpy array* differ from a *list*?
(b) How do we check if two *numpy arrays* are equal or not? 2+2
13. Mention the disadvantages of K-means clustering. What measures can be taken to mitigate those disadvantages? 2+2

Group - C

Answer any *one* question : $8 \times 1 = 8$

14. Explain the fundamental concept of a Linear Support Vector Machine (SVM) for binary classification. Define

the roles of the hyperplane, margins, and support vectors.
How does the SVM determine the 'optimal' decision boundary? 2+2+2+2

15. (a) Discuss the principle behind Decision Tree based classification.

(b) Consider a dataset based on which we will determine whether to play football or not.

Outlook	Temperature	Humidity	Wind	Played football (yes/no)
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Find $E(S)$, $E(S, \text{outlook})$, $IG(S, \text{outlook})$ 2+6

Internal Assessment - 5 marks

P.T.O.

(4)

Paper : COSE405B1

(Pattern Recognition and Image Processing)

Group - A

Answer any *five* questions : 2×5=10

1. What is Pattern Recognition? [CO1]
2. Write the names of any two applications of Pattern Recognition. [CO1]
3. What is a classifier? [CO1]
4. What is K-means clustering? [CO2]
5. Define a digital image. [CO3]
6. What is Histogram Equalization? [CO3]
7. What is image segmentation? [CO4]
8. Write the names of any two morphological operations. [CO4]

Group - B

Answer any *three* questions : 4×3=12

9. Discuss the different components of a Pattern Recognition System. [CO1]
10. Compare and contrast between Supervised and Unsupervised learning. [CO1]
11. Describe K-NN classifier with a simple example. [CO1]

12. How is the quality of a cluster measured? Explain. [CO2]
13. Describe basic image transforms used in image processing. [CO3]

Group - C

Answer any *one* question : $8 \times 1 = 8$

14. (a) What is the need for classification? [CO1]
(b) What is good clustering? [CO2]
(c) What is an outlier? [CO2]
(d) Write short note on Artificial Neural Networks.
 $2+2+(1+3)$ [CO1]
15. (a) Define Digital Image Processing. [CO3]
(b) Explain Histogram Equalization. [CO3]
(c) Describe any two morphological operations.
 $2+3+3$ [CO4]

Internal Assessment - 5 marks

P.T.O.

(6)

Paper : COSE405C1

(Natural Language Processing)

Group - A

Answer any *five* questions : $2 \times 5 = 10$

1. What is NLTK?
2. What is supervised learning?
3. What is Bag-of-word?
4. What is corpus?
5. What is tokenization?
6. What do you mean by text classification?
7. Write down two applications of NLP.
8. Write the significance of CFG is parsing.

Group - B

Answer any *three* questions : $4 \times 3 = 12$

9. How Finite state automata helpful for morphological analysis?
10. Write down the difference between FSA and FST.
11. How N-gram is helpful for statistical calculation of NLP?
12. Briefly explain about Pearson's chi-squared test.

(7)

13. What do you think about mean and variance of a word in NLP?

Group - C

Answer any *one* question : $8 \times 1 = 8$

14. Write short notes on the following topics (any *four*) : $2 \times 4 = 8$

- (a) POS tagger
- (b) Dependency parser
- (c) Hypothesis test
- (d) Inflectional Morphology
- (e) Cognitive Science
- (f) TF-IDF

15. Explain different steps of Text Normalization. Write down the difference between supervised and unsupervised learning in NLP. $4+4$

Internal Assessment - 5 marks
