PG/IIIS/MCA/304/24

MCA 3rd Semester Examination, 2024

MCA

(*Machine Learning*)

PAPER – MCA-304

Full Marks : 100

Time : 3 hours

Answer **all** questions

The figures in the right hand margin indicate marks

Candidates are required to give their answers in their own words as far as practicable

GROUP – A

Answer any **five** questions : 2×5

1. What is overfitting in machine learning, and how can it be prevented ?

2. What is the role of an activation function in a neural network ?

3. What is the difference between bagging and boosting in ensemble methods ?

4. How bias and variance are related to over-fitting and underfitting ?

5. What do you mean by outlier in a dataset ?

6. Explain the purpose of training, validation and test sets.

7. What is the difference between a hard margin and a soft margin in SVM ?

8. Consider an input image of size 12×12 and a 3×3 filter with stride 2 and padding 1. Calculate the size of the output feature map after applying the convolution operation.

GROUP – B

Answer any **four** questions : 15×4

9. Explain the difference among supervised, unsupervised and reinforcement learning. Provide examples of each and discuss their applications. What are the function of numpy, pandas, sklearn, matplotlib and scipy library of python ? $6 + 4 + 5$

10. What is entropy in the context of decision trees, and how is it used to calculate information gain ? Given the following table S, find the entropy(S), entropy(S, Temperature) and information gain(S, Temperature).

Outlook	Temperature	Humidity	Wind	Play Tennis
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

Outlook	Temperature	Humidity	Wind	Play Tennis
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

Explain the concept of ensemble learning in machine learning. (2+2)+(2+4+2)+3

11. Explain the working principle of the Naive Bayes algorithm. How does it use Bayes' Theorem to classify data, and why is the "naive" assumption of independence between features important ? What are the advantages and disadvantages of this algorithm ?

Consider the problem of predicting whether it will rain tomorrow based on the features of the weather today. We have the following data :

Outlook	Temperature	Humidity	Windy	Rain(Class)
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rain	Mild	High	False	Yes
Rain	Cool	Normal	False	Yes

Predict the outcome for a new day with the following features : Outlook = Rain, Temperature = Mild, Humidity = High, Windy = False.

4 + 2 + 2 + 3 + 4

12. What is cross-validation in machine learning, and why is it essential for model selection ? Explain the difference between k-fold cross-validation, leave-one-out cross-validation, and stratified cross-validation. What do you

mean by imbalanced dataset ? How does it affect the performance of a model ? Define feature scaling. 2 + 2 + 6 + 2 + 2 + 1

13. You have trained a machine learning model to predict whether an email is spam (positive class) or not spam (negative class). After evaluating your model on a test dataset, you obtain the following confusion matrix :

	Predicted Spam	Predicted Not Spam
Actual Spam	50	10
Actual Not Spam	5	100

Using this confusion matrix, answer the following questions :

(*i*) What is the accuracy of the model ?

(*ii*) What is the precision of the model for predicting spam (positive class) ?

(*iii*) What is the recall (sensitivity) of the model for predicting spam ?

(*iv*) What is the F1-score of the model for predicting spam ?

(*v*) What is the specificity of the model for predicting not spam (negative class)

How are bias and variance related to overfitting and underfitting of a machine learning model ?

$2 \times 5 + 5$

14. Explain the concept of the hyperplane in Support Vector Machines (SVM). How does the SVM algorithm find the optimal hyperplane for classification, and why is the margin maximized ? Discuss the "one-vs-one" and "one-vs-all" strategies for extending SVM to multi-class classification. $2 + 6 + 2 + 5$

15. What are the potential issues with using a very small or very large value of K in the KNN algorithm ? Explain how the choice of K

influences the model's bias and variance. Why recurrent neural network was introduced and how it works-explain with an example. Explain the difference between linear and logistic regression algorithm.

$2 + 2 + (2 + 5) + 4$

[Internal Assessment – 30 Marks]
