Enrollment No. :
Centre Code :

Invigilator Sign. Date :

Total number of printed pages = 4

2025

(MOOCs) IDC

Paper : MOOCIDC1013

(Essential Mathematics for Machine Learning)

Full Marks : 50

Pass Marks : 20

Time : 2 hours

(Question carries 1 mark each)

Please tick (✓) the correct answer

*Use of correction fluid is prohibited ** Multiple ticks will be treated as incorrect.*

1. **Machine learning models rely heavily on which mathematical field?**
   A. Geography
   B. Calculus
   C. Literature
   D. History
2. **A scalar is:**
   A. A matrix
   B. A single numerical value
   C. A multi-dimensional vector
   D. A tensor of rank 3
3. **A vector of dimension (n) can be represented as:**
   A. A collection of (n) numbers
   B. A single number
   C. A 2D grid
   D. A function
4. **A matrix is a:**
   A. Ordered list
   B. 2D array of numbers
   C. Polynomial
   D. Probability distribution
5. **The dot product of vectors (a) and (b) is:**
   A. (a \times b)
   B. (ab^2)
   C. (a \cdot b)
   D. (a + b)
6. **The identity matrix is:**
   A. Zero matrix
   B. A matrix with 1s on diagonal and 0s elsewhere
   C. A matrix of all 1s
   D. A symmetric matrix
7. **Determinant of a matrix is defined only for:**
   A. Rectangular matrices
   B. Square matrices
   C. Vectors
   D. Tensors
8. **Eigenvalues are associated with:**
   A. Scalars only
   B. Vectors only
   C. Matrices
   D. Polynomials
9. **If (Ax = b), the vector (x) is the:**
   A. Output
   B. Variable
   C. Input
   D. Solution

10. **The derivative of (x^2) is:**

A. (2)
B. (x)
C. (2x)
D. (x^3)

11. **Gradient refers to:**

A. Maximum value
B. Direction of steepest increase
C. Matrix inverse
D. Error minimizer

12. **Gradient descent attempts to:**

A. Maximize cost
B. Minimize cost
C. Randomize parameters
D. Remove data

13. **The learning rate in gradient descent controls:**

A. Number of data points
B. Step size of parameter updates
C. Loss function
D. Gradient direction

14. **A function is convex if it has:**

A. Multiple local minima
B. No minima
C. One global minimum
D. Oscillations

15. **The sigmoid function outputs values between:**

A. (-\infty, \infty)
B. (0,1)
C. (-1,1)
D. (1,10)

16. **The ReLU function is:**

A. (\max(0, x))
B. (x^2)
C. (\frac{1}{1+e^{-x}})
D. (\ln x)

17. **Probability of an event always lies between:**

A. -1 and 1
B. 0 and 1
C. 1 and 10
D. -10 and 10

18. **Mean of numbers measures:**

A. Highest value
B. Central tendency
C. Spread
D. Skewness

19. **Variance measures:**

A. Center
B. Spread of data
C. Minimum
D. Mode

20. **Standard deviation is:**

A. Square of variance
B. Square root of variance
C. Inverse of variance
D. Twice the variance

21. **The covariance of independent variables is:**

A. 0
B. 1
C. -1
D. Undefined

22. **A probability distribution with two parameters:**

A. Matrix
B. Bernoulli
C. Random variable
D. Loss function

23. **A normal distribution is also called:**

A. Uniform distribution
B. Gaussian distribution
C. Poisson distribution
D. Exponential distribution

Contd.

24. The area under a probability density function curve is:

A. 0 B. 1

C. 2 D. Infinite

25. A loss function measures:

A. Accuracy B. Error in prediction

C. Model shape D. Gradient size

26. The common loss function for regression is:

A. Cross-entropy B. Hinge loss

C. Mean squared error D. Log loss

27. Optimization in ML refers to:

A. Speeding up computers B. Minimizing or maximizing objective functions

C. Increasing dataset size D. Randomizing parameters

28. The chain rule is used in:

A. Inverse functions B. Product of numbers

C. Backpropagation D. Data cleaning

29. A tensor is a:

A. Polynomial B. Matrix generalization

C. Random number D. Scalar only

30. A rank-1 tensor is a:

A. Scalar B. Vector

C. Matrix D. Function

31. A rank-2 tensor is a:

A. Scalar B. Vector

C. Matrix D. Graph

32. Correlation measures:

A. Mean value B. Linear relationship

C. Variance D. Matrix rank

33. PCA stands for:

A. Principle Code Analysis B. Principal Component Analysis

C. Partial Component Algorithm D. Parameter Check Algorithm

34. PCA is used for:

A. Increasing dataset size B. Dimensionality reduction

C. Regression D. Feature expansion

35. In ML, overfitting occurs when:

A. Model performs well on all data B. Model performs poorly on training data

C. Model learns noise D. Parameters are fixed

36. Regularization helps to:

A. Increase overfitting B. Reduce overfitting

C. Increase dataset D. Remove labels

37. L1 regularization is also called:

A. Ridge B. Lasso

C. Linear decay D. Softmax

Turn Over

38. L2 regularization is also known as:

A. Lasso
B. Cross-entropy
C. Ridge
D. SVM

39. A hyperplane is:

A. A 1D line
B. A plane in any dimension
C. A point
D. A scalar

40. The softmax function converts inputs into:

A. Random numbers
B. Probabilities
C. Matrices
D. Zeros

41. Conditional probability is written as:

A. (P(A+B))
B. (P(A|B))
C. (P(A \cup B))
D. (AB)

42. Bayes' theorem uses:

A. Derivatives
B. Inverses
C. Conditional probability
D. Linear algebra

43. A function (f(x) = \ln x) is defined for:

A. (x< 0)
B. (x > 0)
C. All x
D. (x = 0)

44. A matrix multiplication (AB) is defined when:

A. Columns of A = Rows of B
B. Rows of A = Rows of B
C. Columns of A = Columns of B
D. Any size works

45. The transpose of a matrix swaps:

A. Rows into columns
B. Numbers
C. Diagonal elements
D. Determinants

46. The rank of a matrix is the number of:

A. Columns
B. Rows
C. Linearly independent rows or columns
D. Diagonal elements

47. If the determinant of a matrix is zero, it is:

A. Invertible
B. Non-invertible
C. Diagonal
D. Symmetric

48. Gradient descent stops when:

A. Loss is minimized
B. Loss increases
C. Data changes
D. Learning rate becomes zero

49. Normalization in ML helps to:

A. Remove errors
B. Scale features
C. Add noise
D. Increase variance

50. A function is differentiable if it is:

A. Continuous
B. Smooth enough to have a derivative
C. Linear
D. Piecewise constant
