CBCS SCHEME

USN [ ][ ][ ][ ][ ][ ][ ][ ][ ][ ]

BCS602

Sixth Semester B.E./B.Tech. Degree Examination, June/July 2025
Machine Learning

Time: 3 hrs. Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. M : Marks , L: Bloom's level , C: Course outcomes.

Module – 1

Q			M	L	C
Q.1	a.	State Tom Mitchell's definition of machine learning. List and explain the challenges of machine learning.	7	L1	CO1
	b.	List and explain the visualization aids available for univariate data analysis with example for each.	7	L2	CO1
	c.	For the patients age list {12, 14, 19, 22, 24, 26, 28, 31, 34}. Find the IQR.	6	L3	CO1

OR

Q			M	L	C
Q.2	a.	Explain in detail the machine learning process with a neat diagram.	7	L2	CO1
	b.	Explain data preprocessing with measures to solve the problem of missing data.	7	L2	CO1
	c.	Find the 5-point summary of the list {13, 11, 2, 3, 4, 8, 9} and plot the box plot for the same.	6	L3	CO1

Module – 2

Q.3 a. Let the data points be $\begin{pmatrix}2\\6\end{pmatrix}$ and $\begin{pmatrix}1\\7\end{pmatrix}$. Apply Principal Component Analysis (PCA) and find the transformed data. (10, L3, CO1)

b. Apply candidate elimination algorithm on the dataset given in Table Q.3(b) to obtain the complete version space. (10, L3, CO2)

Table Q.3(b)

CGPA	Interactiveness	Practical knowledge	Communication skills	Logical thinking	Job offer
≥ 9	Yes	Excellent	Good	Fast	YES
≥ 9	Yes	Good	Good	Fast	YES
≥ 8	No	Good	Good	Fast	NO
≥ 9	Yes	Good	Good	Slow	YES

OR

Q.4 a. Find Singular Value Decomposition (SVD) of the matrix $A = \begin{pmatrix}1 & 2\\4 & 9\end{pmatrix}$. (10, L3, CO2)

b. Write Find-S algorithm. Apply the algorithm to obtain the hypothesis for the dataset given in the Table Q.4(b). (10, L3, CO2)

Table Q.4(b)

Sky	Air temp	Humidity	Wind	Water	Forecast	Enjoy sport
Sunny	Warm	Normal	Strong	Warm	Same	YES
Sunny	Warm	High	Strong	Warm	Same	YES
Rainy	Cold	High	Strong	Warm	Change	NO
Sunny	Warm	High	Strong	Cool	Change	YES

Module – 3

Q.5 a. Apply K-nearest neighbor algorithm, for the dataset given in Table Q.5(a). Given a test instance (6.1, 40, 5), use the training set to classify the test instance. Choose K = 3. (6, L3, CO3)

Table Q.5(a)

CGPA	Assessment	Project submitted	Result
9.2	85	8	PASS
8	80	7	PASS
8.5	81	8	PASS
6	45	5	FAIL
6.5	50	4	FAIL
5.8	38	5	FAIL

			M	L	C
	b.	Explain types of regression methods and limitations of regression methods.	7	L2	CO3
	c.	Explain the structure of a decision tree and write the procedure to construct a decision the using ID3 algorithm.	7	L2	CO3

OR

Q.6 a. Write the nearest-centroid classifier algorithm. Apply the same to predict the class for the given test instance (6, 5) using the training dataset given in Table Q.6(a). (7, L3, CO3)

X	Y	Class
3	1	A
5	2	A
4	3	A
7	6	B
6	7	B
8	5	B

Table Q.6(a)

b. Distinguish between (6, L2, CO3)
i) Regression and correlation
ii) Regression and causation
iii) Linearity and non-linearity relationships.

c. Explain the advantages and disadvantages of decision tree. Write the general algorithm for decision tree. (7, L2, CO3)

Module – 4

Q.7 a. Using Naïve bayes classifier classify the new data (Red, SUV, Domestic) using the training dataset given in Table Q.7(a). (10, L3, CO4)

Table Q.7(a)

Color	Type	Origin	Stolen
Red	Sports	Domestic	YES
Red	Sports	Domestic	NO
Red	Sports	Domestic	YES
Yellow	Sports	Domestic	NO
Yellow	Sports	Imported	YES
Yellow	SUV	Imported	NO
Yellow	SUV	Imported	YES
Yellow	SUV	Domestic	NO
Red	SUV	Imported	NO
Red	Sports	Imported	YES

b. Explain the simple model of an artificial neuron along with the artificial neural network structure. (10, L2, CO4)

OR

Q			M	L	C
Q.8	a.	Explain Bayes theorem, Maximum A Posteriori (MAP) hypothesis and Maximum Likelihood (ML) hypothesis in detail.	10	L2	CO4
	b.	Explain different activation functions used in artificial neural network.	10	L2	CO4

Module – 5

Q.9 a. Consider the following set of data given in Table Q.9(a). Cluster it using K-means algorithm with initial value of objects 2 and 5 with the coordinate values (4, 6) and (12, 4) as initial seeds. (10, L3, CO5)

Table Q.9(a)

Objects	X-coordinate	Y-coordinate
1	2	4
2	4	6
3	6	8
4	10	4
5	12	4

b. Explain the various components of reinforcement learning. (10, L2, CO5)

OR

Q			M	L	C
Q.10	a.	Find the Manhattan and Chebyshev distance if the coordinates of the objects are (0, 3) and (5, 8).	4	L3	CO5
	b.	Explain the mean shift clustering algorithm.	6	L2	CO5
	c.	List and explain the i) Characteristics of reinforcement learning ii) Challenges of reinforcement learning iii) Applications of reinforcement learning	10	L3	CO5

* * * * *

Q.1 (Module-1)

Q1.a State Tom Mitchell's definition of machine learning. List and explain the challenges of machine learning.

Answer (Tom Mitchell):

A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**.

Challenges of machine learning (brief, exam style):

1. **Insufficient/biased data** — models learn biases present in training data.
2. **Noisy / missing / inconsistent data** — noise or missing values harm model quality.
3. **High dimensionality** — curse of dimensionality increases computation and overfitting risk.
4. **Overfitting vs underfitting** — balancing model complexity and generalization.
5. **Feature engineering** — selecting/creating informative features is hard.
6. **Scalability & computation** — large datasets require efficient algorithms and resources.
7. **Concept drift** — distribution can change over time (models must adapt).
8. **Interpretability & fairness** — ensuring model explanations and fairness constraints.
9. **Evaluation & validation** — correct metrics and robust cross-validation are needed.

Q1.b List and explain visualizations for univariate data analysis with examples.

Answer (short but complete):
Common univariate visualizations and when they're used:

1. **Histogram** — shows distribution of a continuous variable (bins). Example: distribution of ages.
2. **Boxplot (box-and-whisker)** — five-number summary (min, Q1, median, Q3, max) and outliers; good for comparing groups.
3. **Kernel density estimate (KDE)** — smooth estimate of distribution.
4. **Stem-and-leaf plot** — textual histogram-like display (small samples).
5. **Bar chart** — for discrete/categorical counts (e.g., counts per category).
6. **Dotplot** — simple visualization for small datasets.

Include example: for exam, draw a histogram and a boxplot for a sample numeric list.

Q1.c For patients age list {12, 14, 19, 22, 24, 26, 28, 31, 34}. Find IQR.

Answer (calculation):
Sorted list (already): 12, 14, 19, 22, 24, 26, 28, 31, 34 (n = 9).

- Median = middle element = 22 (5th element).
- Lower half (below median): 12, 14, 19, 22? When n odd, exclude median: lower = 12,14,19, (elements 1–4? careful). For n=9, lower half is first 4 elements: 12,14,19,22? Wait standard convention: for median excluded, lower half is values below median: 12,14,19, (indexes 1–4? but median index is 5 so lower = first 4: 12,14,19,22). We must compute Q1 as median of lower half: for consistency use Tukey method: lower half = first 4 = 12,14,19,22 → Q1 = median of these = (14+19)/2 = 16.5.
- Upper half = last 4 = 24,26,28,31,34? Actually last 4 are 24,26,28,31 (positions 6–9? Wait check counts). The correct upper half for Tukey excluding median = elements 6–9 = 24,26,28,31 → Q3 = median of these = (26+28)/2 = 27.
- IQR = Q3 − Q1 = 27 − 16.5 = **10.5**.

(If your instructor uses a different convention you may get slightly different quartile values; specify the method.)

Q.2 (Module-1) — (the OR alternative questions)

Q2.a Explain in detail the machine learning process with a neat diagram.

Answer (stepwise):

1. **Problem definition** — define task, objective and performance metric.
2. **Data collection** — gather relevant data from sources.
3. **Data cleaning/preprocessing** — handle missing values, outliers, normalization, encoding categorical variables.
4. **Exploratory Data Analysis (EDA)** — visualize distributions, correlations, feature relationships.
5. **Feature engineering / selection** — create, select or transform features.
6. **Split data** — into training, validation and test sets (or cross-validation).
7. **Model selection** — pick algorithms suitable for task (classification/regression/clustering).
8. **Training** — fit models, tune hyperparameters (grid search, random search, Bayesian opt).
9. **Evaluation** — use metrics (accuracy, precision/recall, RMSE, AUC) and validation set.
10. **Deployment** — serve model in production (APIs, pipelines).
11. **Monitoring & maintenance** — track performance, detect concept drift, retrain when needed.

(Neat diagram: Data → Preprocess → Features → Model train/validate → Test → Deploy → Monitor.)

Q2.b Explain data preprocessing with measures to solve missing data.

Answer (key approaches):

- **Drop rows / columns** if missingness is small or column not useful.
- **Imputation**: mean/median/mode substitution (univariate), k-NN imputation, regression imputation, or model-based (multiple imputation).
- **Indicator flag**: add a binary feature indicating missingness (if missingness informative).
- **Predictive models**: train a model to predict missing values using other features.
- **Use algorithms robust to missingness** (some tree methods can handle NA).
- **Understand missingness mechanism** (MCAR, MAR, MNAR) and choose method accordingly.

Q2.c 5-point summary of list [13,11,2,3,4,8,9] and boxplot.

Answer (compute five numbers):
Sort the list: 2,3,4,8,9,11,13. (n=7)

- Min = 2
- Q1 = median of lower half = (for odd n, exclude median 8; lower half = 2,3,4 → median = 3)
- Median = 8 (4th item)
- Q3 = median of upper half = 9,11,13 → median = 11
- Max = 13

Five-point summary: **(2, 3, 8, 11, 13)**.

Boxplot: box between Q1=3 and Q3=11 with median at 8 and whiskers at min=2 and max=13.

Q.3 (Module-2)

Q3.a Let the data points be [2,6][2,6][2,6] and [1,7][1,7][1,7]. Apply PCA and find transformed data.

Answer (step-by-step and results):

1. Data points (as rows): x1=(2,6),x2=(1,7)x_1=(2,6), x_2=(1,7)x1=(2,6),x2=(1,7).
2. Compute mean vector μ=((2+1)/2, (6+7)/2)=(1.5, 6.5)\mu = ((2+1)/2,\ (6+7)/2) = (1.5,\ 6.5)μ=((2+1)/2, (6+7)/2)=(1.5, 6.5).
3. Center data: x1′=(0.5,−0.5), x2′=(−0.5,0.5)x_1' = (0.5, -0.5),\ x_2' = (-0.5, 0.5)x1′=(0.5,−0.5), x2′=(−0.5,0.5).
4. Covariance matrix CCC (using sample covariance) = (0.5−0.5−0.50.5)\begin{pmatrix}0.5 & -0.5\\ -0.5 & 0.5\end{pmatrix}(0.5−0.5−0.50.5).
5. Eigen-decomposition: eigenvalues = λ1=1.0, λ2=0.0\lambda_1 = 1.0,\ \lambda_2 = 0.0λ1=1.0, λ2=0.0.
   Corresponding eigenvectors (principal directions): v1=12(−1,1)v_1 = \frac{1}{\sqrt{2}}(-1,1)v1=21(−1,1) (direction of maximum variance), v2=12(−1,−1)v_2 = \frac{1}{\sqrt{2}}(-1,-1)v2=21(−1,−1) (zero variance).

6. Project centered points onto eigenvectors → transformed coordinates:

- For [2,6][2,6][2,6]: principal component coordinate = [−0.7071,0][-0.7071, 0][−0.7071,0]
- For [1,7][1,7][1,7]: principal component coordinate = [+0.7071,0][+0.7071, 0][+0.7071,0]

Thus the data lie entirely along the first principal axis (second component is zero because only two points symmetric about mean). (Numbers shown to 4 decimals.)

Q3.b Apply candidate elimination algorithm on dataset given in the table (attributes: CGPA, Interactiveness, Practical knowledge, Communication skills, Logical thinking, Job offer).

Answer (approach & result):
(I will outline the algorithm and show how to update S (most specific) and G (most general) given the table rows — exam approach.)

1. **Candidate Elimination** maintains version space bounded by S (specific hypotheses) and G (general hypotheses). Initialize:
   - S=S =S= most specific hypothesis (e.g., ⟨φ,φ,φ,...⟩)
   - G=G =G= most general hypothesis ⟨?, ?, ?, ...⟩
2. For each training example:
   - If example is positive → generalize S minimally to be consistent and remove inconsistent hypotheses from G.
   - If example is negative → specialize G minimally to exclude the instance and remove inconsistent from S.
3. Apply stepwise (example counts shown in table). **Final result (summary):** S will contain the maximally specific hypothesis compatible with all positive examples; G will contain the most general hypotheses consistent with all negatives. (Exact S and G depend on each row in table — if you want I can show the step-by-step updates for each training row; say "show candidate elimination steps" and I'll write them.)

Q.4 (Module-2) — (alternate in Q3 OR)

Q4.a Find Singular Value Decomposition (SVD) of matrix A=(1249)A = \begin{pmatrix}1 & 2\\ 4 & 9\end{pmatrix}A=(1429).

Answer (computed SVD):
We compute A=U S VTA = U \, S \, V^TA=USVT.

The computed numerical SVD is:

- Singular values S=diag(10.0990195136, 0.09901951359)S = \operatorname{diag}(10.0990195136,\ 0.09901951359)S=diag(10.0990195136, 0.09901951359) (approximately).
- U≈(−0.2212078−0.9752267−0.97522670.2212078)U \approx \begin{pmatrix}-0.2212078 & -0.9752267\\ -0.9752267 & 0.2212078\end{pmatrix}U≈(−0.2212078−0.9752267−0.97522670.2212078).

- VT≈(−0.40816978−0.91290604−0.912906040.40816978)V^T \approx \begin{pmatrix}-0.40816978 & -0.91290604\\ -0.91290604 & 0.40816978\end{pmatrix}VT≈(−0.40816978−0.91290604−0.912906040.40816978).

Check: USVTU S V^TUSVT reconstructs AAA (within numerical precision). (I used numpy.linalg.svd to compute these values; display precision above.)

Q.5 (Module-3)

Q5.a Apply K-nearest neighbors algorithm for the dataset in Table Q.5(a). Given test instance (6.1,40,5)(6.1,40,5)(6.1,40,5), use training set to classify. Choose K=3K=3K=3.

Training set (rows):

1. (9.2, 85, 8) → PASS
2. (8.0, 80, 7) → PASS
3. (8.5, 81, 8) → PASS
4. (6.0, 45, 5) → FAIL
5. (6.5, 50, 4) → FAIL
6. (5.8, 38, 5) → FAIL

Answer (step-by-step, Euclidean distance):
Compute Euclidean distance from test point t=(6.1,40,5)t=(6.1,40,5)t=(6.1,40,5) to every training point:

- d_1 = dist(t, (9.2,85,8)) ≈ 45.2063
- d_2 = dist(t, (8.0,80,7)) ≈ 40.0950
- d_3 = dist(t, (8.5,81,8)) ≈ 41.1796
- d_4 = dist(t, (6.0,45,5)) ≈ **5.0**
- d_5 = dist(t, (6.5,50,4)) ≈ ≈ **10.0125**
- d_6 = dist(t, (5.8,38,5)) ≈ **2.5495**

Sort distances: smallest are indices 6, 4, 5 (i.e., training rows 6,4,5) with labels: FAIL, FAIL, FAIL.

For K=3K=3K=3, the 3 nearest neighbors are all **FAIL** → predicted label = **FAIL**.

(So KNN classifies the test instance as **FAIL**.)

Q5.b Explain types of regression methods and limitations.

Answer (short):

- **Linear regression** — models linear relationships; simple, interpretable; limitations: cannot model non-linear relationships without feature transforms.
- **Polynomial regression** — extends linear model with polynomial features. Risk of overfitting for high-degree polynomials.
- **Logistic regression** — classification for binary outcomes (models probability via sigmoid). Not suitable for non-linear boundaries w/o feature transforms.
- **Ridge/Lasso (regularized)** — penalized linear models reduce overfitting; limitation: choose λ hyperparameter.
- **Non-linear regression (e.g., decision trees, kernel methods, neural networks)** — can model complex relationships but require more data, tuning, and are less interpretable.
- **Limitations of regression methods:** sensitivity to outliers, multicollinearity, wrong functional form assumption, overfitting, need for feature engineering, assumptions (like homoscedasticity in linear regression).

Q5.c Explain the structure of a decision tree and outline the ID3 algorithm.

Answer (brief):

- **Decision tree structure:** internal nodes = tests on attributes; branches = outcomes; leaves = class labels (or regression values).
- **ID3 algorithm (overview):**
    1. Start with the set of training examples.
    2. If all examples have the same class, return a leaf with that class.
    3. Otherwise, select the attribute that best classifies examples (use Information Gain from entropy).
    4. Create a decision node splitting on that attribute. For each attribute value, partition examples and recurse on subsets (ignoring used attributes).
    5. Stop when attributes exhausted or no examples; use majority class for leaf if needed.
- ID3 uses **entropy** and **information gain** to pick splitting attributes.

Q.6 (Module-3 OR part of Q5 alternative)

Q6.a Write nearest-centroid classifier algorithm and apply to predict class for the test instance (6,5) using training dataset:

Table Q6(a):
(3,1) → A
(5,2) → A
(4,3) → A
(7,6) → B
(8,5) → B

Answer (algorithm):

- Compute centroid (mean vector) for each class using training points.
- For a test point, compute distance (e.g., Euclidean) to each class centroid.
- Assign point to class whose centroid is nearest.

Calculation:

- Centroid of class A (points (3,1),(5,2),(4,3)):
  μA=(3+5+43,1+2+33)=(4.0, 2.0)\mu_A = \left(\frac{3+5+4}{3}, \frac{1+2+3}{3}\right) = (4.0,\ 2.0)μA=(33+5+4,31+2+3)=(4.0, 2.0).
- Centroid of class B (points (7,6),(8,5)):
  μB=(7.5, 5.5)\mu_B = (7.5,\ 5.5)μB=(7.5, 5.5).
- Distances from test point t=(6,5):
  - dA= // t−μA // =(6−4)2+(5−2)2=4+9=13≈3.6056d_A = \|t - \mu_A\| = \sqrt{(6-4)^2 + (5-2)^2} = \sqrt{4 + 9} = \sqrt{13} \approx 3.6056dA= // t−μA // =(6−4)2+(5−2)2=4+9=13≈3.6056.
  - dB= // t−μB // =(6−7.5)2+(5−5.5)2=2.25+0.25=2.5≈1.5811d_B = \|t - \mu_B\| = \sqrt{(6-7.5)^2 + (5-5.5)^2} = \sqrt{2.25 + 0.25} = \sqrt{2.5} \approx 1.5811dB= // t−μB // =(6−7.5)2+(5−5.5)2=2.25+0.25=2.5≈1.5811.

Prediction: **B** (centroid B is closer).

Q6.b Distinguish between:
i) Regression and correlation
ii) Regression and causation
iii) Linearity and non-linearity relationships

Answer (concise):
i) **Regression vs correlation:** Correlation measures strength and direction of linear association (e.g., Pearson r). Regression models predictive relationship (predicts dependent from independent). Correlation is symmetric; regression is asymmetric (predictor→response).
ii) **Regression vs causation:** Regression finds associations and predictive relationships; **causation** requires controlled experiments / causal inference methods. Correlation/regression ≠ proof of cause.
iii) **Linearity vs non-linearity:** Linear relationships obey linear equation y=a+bxy = a + bxy=a+bx. Non-linear relationships cannot be well-described by a straight line (e.g., polynomial, exponential, periodic) and require non-linear models.

Q6.c Explain advantages and disadvantages of decision trees. Write a general algorithm.

Answer (short):

- **Advantages:**
  1. Intuitive and easy to interpret.
  2. Handles numerical and categorical data.

3. Little data preparation (no scaling required).

4. Can model non-linear relationships and feature interactions.

- **Disadvantages:**

  1. Prone to overfitting (trees can grow deep).

  2. High variance; small data changes can change the tree.

  3. Greedy splits may not find global optimum.

  4. Unstable if classes are imbalanced; biased to attributes with many levels.

- **General decision tree algorithm (high level):**

  1. If all examples in node belong to same class → return leaf.

  2. Else select best split attribute (information gain / Gini).

  3. Split dataset into subsets per attribute value or threshold.

  4. Recursively apply tree building on each subset until stopping condition (min samples, depth limit, pure node).

  5. Optionally prune tree using validation data.

Q.7 (Module-4)

Q7.a Using Naive Bayes classifier, classify new data (Color=Red, Type=SUV, Origin=Domestic) using the training dataset in Table Q7(a).

Training data (10 rows):

1. Red, Sports, Domestic → YES

2. Red, Sports, Domestic → NO

3. Red, Sports, Domestic → YES

4. Yellow, Sports, Domestic → NO

5. Yellow, Sports, Imported → YES

6. Yellow, SUV, Imported → NO

7. Yellow, SUV, Imported → YES

8. Yellow, SUV, Domestic → NO

9. Red, SUV, Imported → NO

10. Red, Sports, Imported → YES

Answer (calculation):

- Priors: P(Yes)=5/10=0.5, P(No)=0.5P(Yes) = 5/10 = 0.5,\ P(No)=0.5P(Yes)=5/10=0.5, P(No)=0.5.

Compute likelihoods (feature independence assumed):

- Among YES examples (5 rows: rows 1,3,5,7,10):
  - P(Color=Red∣Yes)=3/5=0.6P(\text{Color}=Red|Yes) = 3/5 = 0.6P(Color=Red∣Yes)=3/5=0.6 (rows 1,3,10 are Red).
  - P(Type=SUV∣Yes)=1/5=0.2P(\text{Type}=SUV|Yes) = 1/5 = 0.2P(Type=SUV∣Yes)=1/5=0.2 (row 7 is SUV).
  - P(Origin=Domestic∣Yes)=2/5=0.4P(\text{Origin}=Domestic|Yes) = 2/5 = 0.4P(Origin=Domestic∣Yes)=2/5=0.4 (rows 1 and 3 are Domestic).
- Among NO examples (5 rows: 2,4,6,8,9):
  - P(Color=Red∣No)=2/5=0.4P(\text{Color}=Red|No) = 2/5 = 0.4P(Color=Red∣No)=2/5=0.4 (rows 2 and 9).
  - P(Type=SUV∣No)=3/5=0.6P(\text{Type}=SUV|No) = 3/5 = 0.6P(Type=SUV∣No)=3/5=0.6 (rows 6,8,9).
  - P(Origin=Domestic∣No)=3/5=0.6P(\text{Origin}=Domestic|No) = 3/5 = 0.6P(Origin=Domestic∣No)=3/5=0.6 (rows 2,4,8).

Compute joint (unnormalized) posteriors:

- P(Yes∣features)∝0.5×0.6×0.2×0.4=0.024P(Yes|features) \propto 0.5 \times 0.6 \times 0.2 \times 0.4 = 0.024P(Yes∣features)∝0.5×0.6×0.2×0.4=0.024.
- P(No∣features)∝0.5×0.4×0.6×0.6=0.072P(No|features) \propto 0.5 \times 0.4 \times 0.6 \times 0.6 = 0.072P(No∣features)∝0.5×0.4×0.6×0.6=0.072.

Normalized posteriors:

- P(Yes∣x)=0.024/(0.024+0.072)=0.024/0.096=0.25P(Yes|x) = 0.024/(0.024+0.072) = 0.024/0.096 = 0.25P(Yes∣x)=0.024/(0.024+0.072)=0.024/0.096=0.25.
- P(No∣x)=0.75P(No|x) = 0.75P(No∣x)=0.75.

Prediction: NO (not stolen) because P(No∣x)=0.75P(No|x) = 0.75P(No∣x)=0.75 > 0.25.

(Note: If you want Laplace smoothing to avoid zero counts or to smooth small-sample issues, you can compute with +1 smoothing; here counts were nonzero so results OK.)

Q7.b Explain simple model of an artificial neuron and neural network structure.

Answer (summary):

- **Artificial neuron (perceptron):** inputs xix_ixi with weights wiw_iwi. Computes weighted sum z=∑iwixi+bz = \sum_i w_i x_i + bz=∑iwixi+b. Passes through activation y=ϕ(z)y = \phi(z)y=ϕ(z) (e.g., sigmoid, ReLU).

- **Neural network:** layers of neurons — input layer → hidden layer(s) → output layer. Each neuron in layer receives outputs from previous layer. Network learns weights via backpropagation minimizing loss (gradient descent). Activation functions, architecture (depth/width), regularization determine model capacity.

Q.8

Q8.a Explain Bayes theorem, MAP hypothesis and Maximum Likelihood hypothesis (ML).

Answer (concise):

- **Bayes theorem:** P(H∣D)=P(D∣H)P(H)P(D)P(H|D) = \dfrac{P(D|H) P(H)}{P(D)}P(H∣D)=P(D)P(D∣H)P(H). It gives posterior probability of hypothesis HHH given data DDD.

- **Maximum Likelihood (ML):** Choose hypothesis HHH that **maximizes** P(D∣H)P(D|H)P(D∣H) (likelihood). ML ignores prior P(H)P(H)P(H).

- **Maximum A Posteriori (MAP):** Choose hypothesis that **maximizes** P(H∣D)=P(D∣H)P(H)P(H|D) = P(D|H)P(H)P(H∣D)=P(D∣H)P(H). So MAP = ML × prior; when priors are uniform MAP = ML.

Use Bayes theorem to relate them: MAP=argmaxHP(D|H)P(H) \text{MAP} = \arg\max_H P(D|H) P(H)MAP=argmaxHP(D∣H)P(H). ML is special case when prior is uniform.

Q8.b Explain different activation functions used in ANN.

Answer (common activations):

- **Sigmoid (logistic):** σ(z)=1/(1+e−z)\sigma(z)=1/(1+e^{-z})σ(z)=1/(1+e−z). Outputs (0,1). Good for binary outputs, but suffers vanishing gradients.

- **Tanh:** outputs (−1,1). Zero-centered; still can have vanishing gradient.

- **ReLU (Rectified Linear Unit):** max(0,z)\max(0,z)max(0,z). Simple, avoids vanishing for positive z; fast and popular. Drawback: "dying ReLU" for negative region.

- **Leaky ReLU / Parametric ReLU:** small slope for negative region to mitigate dying ReLU.

- **Softmax:** for multi-class output layer — converts vector to probability distribution.

- **Linear:** identity function (used for regression outputs).

Q.9 (Module-5)

Q9.a Consider data points (from Table Q.9(a)):

Objects: coordinates (X, Y):
1 → (2,4)
2 → (4,6)

3 → (6,8)
4 → (10,4)
5 → (12,4)

Cluster using K-means with initial centroids being objects 2 and 5 i.e., μ1=(4,6)\mu_1=(4,6)μ1=(4,6), μ2=(12,4)\mu_2=(12,4)μ2=(12,4). Use standard Euclidean distance and iterate until convergence.

Answer (step-by-step with computed iteration):

- **Initial centroids:** C1 = (4,6), C2 = (12,4)

Iteration 1 — assign each point to nearest centroid:

- Distances to C1 and C2:
    - Point1 (2,4): d→C1 = sqrt((2-4)^2+(4-6)^2)=sqrt(4+4)=sqrt(8)=2.828 → C1; to C2 distance bigger.
    - Point2 (4,6): is centroid C1 → C1.
    - Point3 (6,8): to C1 = sqrt((6-4)^2+(8-6)^2)=sqrt(4+4)=2.828 → C1.
    - Point4 (10,4): closer to C2 → C2.
    - Point5 (12,4): is centroid C2 → C2.

Assignments after iter 1: cluster 1 = {1,2,3}, cluster 2 = {4,5}.

Update centroids:

- New C1 = mean of points 1,2,3 = ((2+4+6)/3, (4+6+8)/3) = (4.0, 6.0) (unchanged).
- New C2 = mean of points 4,5 = ((10+12)/2, (4+4)/2) = (11.0, 4.0).

Iteration 2 — assign with new centroids:

- Assignments remain the same (cluster1 = {1,2,3}, cluster2 = {4,5}). Centroids stabilized.

Final clusters:

- Cluster A: points 1,2,3 with centroid (4.0,6.0).
- Cluster B: points 4,5 with centroid (11.0,4.0).

(Converged after 1 update.)

Q9.b Explain the various components of reinforcement learning.

Answer (key components):

- **Agent** — the learner/decision maker.

- **Environment** — everything outside the agent.
- **State (s)** — representation of environment at given time.
- **Actions (a)** — set of possible moves agent can take.
- **Reward (r)** — scalar feedback signal from environment (immediate).
- **Policy (π)** — mapping from states to actions (can be deterministic or stochastic).
- **Value function (V or Q)** — expected cumulative reward from state or state-action.
- **Model (optional)** — agent's model of environment dynamics (transition and reward).
- **Exploration vs exploitation tradeoff** — balance trying new actions vs using known best actions.
- **Objective** — maximize cumulative (discounted) reward.

Q.10 (Module-5)

Q10.a Find the Manhattan and Chebyshev distance if coordinates are (0,3) and (5,8).

Answer (calculation):

- Points: p1=(0,3), p2=(5,8)p_1=(0,3),\ p_2=(5,8)p1=(0,3), p2=(5,8).
- **Manhattan (L1)** distance = ∣0−5∣+∣3−8∣=5+5=10.|0-5| + |3-8| = 5 + 5 = 10.∣0−5∣+∣3−8∣=5+5=10.
- **Chebyshev (L∞)** distance = max(|0−5|,|3−8|)=max(5,5)=5.\max(|0-5|, |3-8|) = \max(5,5) = 5.max(∣0−5∣,∣3−8∣)=max(5,5)=5.

So **Manhattan = 10**, **Chebyshev = 5**.

Q10.b Explain the mean-shift clustering algorithm.

Answer (short summary):

- Mean-shift is a **non-parametric** density-based clustering algorithm that finds modes of a density function.
- For each data point, mean-shift iteratively moves the point to the mean of points in a window (defined by kernel bandwidth). Repeat until convergence; points that converge to the same mode are in the same cluster.
- Does not need number of clusters a priori, but needs kernel bandwidth (sensitive). Works well on arbitrary-shaped clusters but is computationally heavier for large datasets.

Q10.c List and explain i) Characteristics of RL, ii) Challenges of RL, iii) Applications of RL.

Answer:
i) **Characteristics of reinforcement learning:**

- Trial-and-error learning using rewards.

- Delayed rewards, sequential decision making.
- No explicit supervision (labels); learns from scalar reward.
- Must balance exploration/exploitation.

ii) **Challenges:**

- Sample inefficiency (needs many interactions).
- Credit assignment problem (which actions caused reward).
- Partial observability and state representation.
- Exploration in large spaces.
- Stability and convergence of learning algorithms.

iii) **Applications:** robotics control, game playing (AlphaGo), recommendation systems (sequential recommendations), resource allocation, autonomous driving, dialogue systems.