Machine Learning

Course Code: CSE473

Unit-wise Assignment Question Bank

Units 1 to 5 · 4, 8, 10 and 12 Mark Questions

Programme	B.Tech. — Computer Science & Engineering
Department	Department of Computer Science & Engineering
School	School of Computing Science & Engineering
Institution	Sharda University, Greater Noida
Faculty	Dr. Gopal Chandra Jana
Total Questions	80

Course page: `https://www.gcjana.in/courses/shardauniversity/2601/machine-learning/`

Post your doubts: gcjana.in/.../machine-learning/#Post-doubts

Every unit contains at least eight numerical problems. A complete answer key is provided in the Appendix.

General Instructions to Students

1. All assignments are to be **handwritten** unless stated otherwise, and submitted as a **single PDF** through the course page.
2. For every numerical question, **state the formula first**, then substitute, then simplify. A bare final answer earns **at most 40%** of the marks.
3. Round final answers to **four decimal places** unless the question says otherwise. Intermediate values should not be rounded prematurely.
4. Standard statistical table values (z, t, χ^2) are supplied inside the questions wherever needed.
5. Diagrams must be **neatly labelled**; unlabelled diagrams earn no marks.
6. Assignment 1 covers **Units 1 and 2**; Assignment 2 covers **Units 3, 4 and 5**.
7. Copied or identical submissions will be awarded **zero** for all parties involved.
8. Write the **question number and marks** clearly before each answer.

Suggested answer length:

Marks	Expected length	Expected content
4	∼half page	definition/statement + one example or one calculation
8	∼1 page	explanation + diagram *or* complete numerical
10	∼1.5 pages	full explanation + labelled diagram + numerical
12	2–2.5 pages	derivation + diagram + numerical + discussion

1 Unit 1 — Core Concepts of Machine Learning

Unit 1: Core Concepts of Machine Learning

A. What is ML; problem framing; applications; standard learning tasks. **B.** Learning stages; learning scenarios; generalization. **C.** Data preparation, preprocessing, cleaning, integration, reduction, transformation and discretization.

SECTION A — 4 Mark Questions (Attempt all six)

Q1. [4 Marks]
State *Mitchell's* definition of machine learning. Write the triple $\langle T, P, E\rangle$ for (a) a spam filter and (b) a handwriting recogniser.

Q2. [4 Marks]
Classify each of the following as classification, regression, ranking, clustering or dimensionality reduction, giving a one-line reason: (i) predicting tomorrow's rainfall in mm; (ii) grouping unlabelled news articles; (iii) ordering search results for a query; (iv) compressing 784 pixel values into 2 numbers.

Q3. [4 Marks]
An attribute *marks* has min = 20, max = 120, mean = 60 and standard deviation = 25. Normalize the value $v = 45$ by (a) min–max to $[0, 1]$, and (b) z-score. Show both formulas.

Q4. [4 Marks]
A model predicts $\hat{y} = (11, 11, 16, 17)$ while the true values are $y = (10, 12, 15, 18)$. Compute the MSE, RMSE and MAE.

Q5. [4 Marks]
Out of 2000 transactions, bread appears in 600, jam in 400, and both together in 240. Compute the *support*, *confidence* and *lift* of the rule bread $\Rightarrow$ jam, and state whether the association is positive.

Q6. [4 Marks]
A dataset has 8 attributes. How many possible attribute subsets exist? Repeat for 15 attributes, and state in one line why exhaustive attribute subset selection is infeasible.

SECTION B — 8 Mark Questions (Attempt all four)

Q7. [8 Marks]
Differentiate between *supervised*, *unsupervised* and *reinforcement* learning on any six points in tabular form. For each, name two algorithms and one real application.

Q8. [8 Marks]
Explain *parameters* versus *hyperparameters* with three examples of each. Explain the roles of the training, validation and test sets, and state clearly why the test set must never be used for tuning.

Q9. [8 Marks]
Sorted price data: 5, 10, 11, 13, 15, 35, 50, 55, 72.

(a) Partition into three bins by *equal frequency* and smooth by *bin means*.

(b) Smooth the same bins by *bin boundaries*.

(c) Partition the data by *equal width* into three bins and list the members of each.

(d) State one advantage of each partitioning scheme.

Q10. [8 Marks]
For the two numeric attributes $A = (1, 2, 3, 4, 5)$ and $B = (2, 4, 5, 4, 5)$, compute the covariance and the Pearson correlation coefficient. Interpret the sign and magnitude, and state whether either

attribute is redundant. [Use $\sigma_A = 1.4142$, $\sigma_B = 1.0954$, $\sum a_i b_i = 66$.]

SECTION C — 10 Mark Questions (Attempt all three)

Q11. **[10 Marks]**

(a) Draw and explain the complete *machine learning pipeline* with all its stages. (5)

(b) State where most project effort is actually spent and justify the claim "*garbage in, garbage out*". (2)

(c) Explain *data leakage* with one concrete example and one way to detect it. (3)

Q12. **[10 Marks]**

A 2×2 contingency table for *gender* versus *product preference* is:

	Product X	Product Y	Total
Male	60	40	100
Female	20	80	100
Total	80	120	200

(a) Compute the expected count of every cell. (4)

(b) Compute the χ^2 statistic showing all four contributions. (4)

(c) With $\chi^2_{0.05,1} = 3.841$, state your conclusion about correlation. (2)

Q13. **[10 Marks]**

Explain the *four major tasks* of data preprocessing. For *data cleaning*, describe all six standard strategies for handling missing values, and apply mean imputation to the income data 22, 28, ?, 34, 36 (in thousands), stating why the median might be preferred.

SECTION D — 12 Mark Questions (Attempt all three)

Q14. **[12 Marks]**

(a) Explain *overfitting* and *underfitting*, and sketch the error-versus-complexity curve. (4)

(b) State the bias–variance decomposition and explain each term. (4)

(c) A model scores 98% on training data and 61% on test data. Diagnose the problem and give four remedies. (4)

Q15. **[12 Marks]**

(a) Explain all seven *learning scenarios* (supervised, unsupervised, semi-supervised, transductive, on-line, reinforcement, active), stating the data given and the feedback available in each. (7)

(b) Distinguish inductive from transductive inference. (2)

(c) A fraud dataset has 980 genuine and 20 fraudulent records out of 1000. Compute the accuracy of an "always genuine" classifier and explain why the metric is inappropriate. (3)

Q16. **[12 Marks]**

(a) Explain the *data reduction* strategies: dimensionality reduction, numerosity reduction and data-cube aggregation. (4)

(b) Explain *attribute subset selection* and describe forward selection and backward elimination with a diagram. (4)

(c) Explain the four sampling methods (SRSWOR, SRSWR, cluster, stratified) and state which is best for skewed data, with reasons. (4)

2 Unit 2 — Supervised Learning Algorithms, Part One

Unit 2: Supervised Learning Algorithms — Part One

A. How supervised learning works; bias–variance; complexity, dimensionality, noise. **B.** Linear regression, OLS, gradient descent, regularization, prediction. **C.** Logistic regression; model fitting; linear discriminant analysis; applications.

SECTION A — 4 Mark Questions (Attempt all six)

Q1. **[4 Marks]**
An estimator has bias = 3 and variance = 4. Compute its MSE. Repeat for an unbiased estimator with variance 9, and state which you would prefer under squared-error loss.

Q2. **[4 Marks]**
To cover the unit hypercube $[0,1]^n$ with a grid of spacing 0.1, how many points are needed for $n = 1$, $n = 3$ and $n = 6$? Name the phenomenon and state one remedy.

Q3. **[4 Marks]**
For the data $x = (1, 2, 3, 4, 5)$, $y = (3, 4, 6, 7, 9)$ it is given that $S_{xy} = 15$ and $S_{xx} = 10$. Compute the slope θ_1 and intercept θ_0 of the least-squares line.

Q4. **[4 Marks]**
House sizes are 1200, 1800, 900, 1500, 1600 sq.ft. Scale the value 1500 by (a) min–max, (b) mean normalization, and (c) z-score. [Use $\mu = 1400$, $\sigma = 316.2278$.]

Q5. **[4 Marks]**
Compute $\sigma(z)$ for $z = -1,\ 0,\ 1$. Then for $p = 0.8$ compute the odds and the logit, showing the formula for each.

Q6. **[4 Marks]**
In a logistic model $\ln \frac{p}{1-p} = -3 + 0.06x$. Compute the decision boundary (the value of x at which $p = 0.5$) and the odds ratio for a one-unit increase in x. [$e^{0.06} = 1.0618$.]

SECTION B — 8 Mark Questions (Attempt all four)

Q7. **[8 Marks]**
For $x = (1, 2, 3, 4, 5)$ and $y = (3, 4, 6, 7, 9)$ with fitted line $\hat{y} = 1.3 + 1.5x$:

(a) Compute all five predictions and residuals, and verify that $\sum e_i = 0$.

(b) Compute SSE, SST and R^2.

(c) Compute the MSE and interpret the value of R^2.

Q8. **[8 Marks]**
Derive the *normal equation* $\boldsymbol{\theta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$ from the cost function. For the data of the previous question it is given that $\mathbf{X}^\top\mathbf{X} = \begin{pmatrix} 5 & 15 \\ 15 & 55 \end{pmatrix}$ and $\mathbf{X}^\top\mathbf{y} = \begin{pmatrix} 29 \\ 102 \end{pmatrix}$. Compute the determinant, the inverse and hence $\boldsymbol{\theta}$, and verify it against Question 7.

Q9. **[8 Marks]**
Compare *Ridge* and *Lasso* regression on any five points, writing both cost functions. For a no-intercept model with $\sum x_i y_i = 30$ and $\sum x_i^2 = 15$, compute the ridge estimate $\theta = \frac{\sum x_i y_i}{\sum x_i^2 + \lambda}$ for $\lambda = 0,\ 3$ and 15, and comment on the shrinkage.

Q10. **[8 Marks]**
Explain the terms *odds*, *odds ratio* and *logit*, and derive the logit as the inverse of the logistic function. In a clinical trial, 45 of 60 treated patients recovered while 20 of 80 control patients recovered. Compute the odds in each group and the odds ratio, and interpret the result.

SECTION C — 10 Mark Questions (Attempt all three)

Q11. **[10 Marks]**

(a) Derive the gradient-descent update rule for linear regression from the cost $J(\boldsymbol{\theta}) = \frac{1}{2m}\sum(h_{\boldsymbol{\theta}}(x^{(i)}) - y^{(i)})^2$. (4)

(b) For $x = (1, 2, 3, 4, 5)$, $y = (3, 4, 6, 7, 9)$, starting from $\theta_0 = \theta_1 = 0$ with $\alpha = 0.05$, perform *two* iterations, computing the gradients, the updated parameters and J at each step. (6)

Q12. **[10 Marks]**

(a) Explain why squared error is not used as the cost function for logistic regression, and write the log-loss (cross-entropy) function. (4)

(b) For $y = (1, 1, 0, 0)$ and predicted $h = (0.8, 0.6, 0.3, 0.1)$, compute the loss of each sample and the average log loss. (4)

(c) Explain why a confidently wrong prediction is penalised so heavily. (2)

Q13. **[10 Marks]**

Two classes have the observations Class 1: $\{1, 2, 3, 4\}$ and Class 2: $\{6, 7, 8, 9\}$, with equal priors.

(a) Compute the class means and the pooled variance s_p^2. (4)

(b) Compute the LDA decision threshold and classify the points $x = 4.5$ and $x = 5.5$. (4)

(c) Compute the Mahalanobis effect size D^2 and comment on the separability. (2)

SECTION D — 12 Mark Questions (Attempt all three)

Q14. **[12 Marks]**

(a) State and explain the bias–variance decomposition $\mathbb{E}[(y - \hat{f})^2] = \mathrm{Bias}^2 + \mathrm{Var} + \sigma^2$. (4)

(b) Explain how function complexity, training-set size, input dimensionality and output noise each affect the choice of a supervised learning algorithm. (5)

(c) Explain heterogeneity, redundancy and interactions in data, and state how each is handled. (3)

Q15. **[12 Marks]**

(a) Explain *feature scaling* and sketch the cost contours before and after scaling. (4)

(b) Explain how the learning rate α affects convergence, sketching J versus iterations for a good α, one too small and one too large. (4)

(c) Compare gradient descent with the normal equation on any four points, and state when each should be used. (4)

Q16. **[12 Marks]**

(a) Explain the logistic model, including the latent-variable interpretation. (4)

(b) For $\ln \frac{p}{1-p} = -3 + 0.06x$, compute p and the odds at $x = 30$, 50 and 70. (4)

(c) Compare LDA with logistic regression on any four points, and state two practical applications of LDA. (4)

3 Unit 3 — Supervised Learning Algorithms, Part Two

Unit 3: Supervised Learning Algorithms — Part Two

A. Support Vector Machines: linear, non-linear, primal/dual, kernels, ERM, properties. **B.** Artificial Neural Networks; network training; backpropagation. **C.** Decision tree learning (ID3); random forests.

SECTION A — 4 Mark Questions (Attempt all six)

Q1. [4 Marks]
For an SVM with $\mathbf{w} = (3, 4)$, compute $\|\mathbf{w}\|$ and the margin width $2/\|\mathbf{w}\|$. Repeat for $\mathbf{w} = (0.6, 0.8)$.

Q2. [4 Marks]
Compute the hinge loss $L = \max(0, 1 - yf(\mathbf{x}))$ for $(y, f) = (+1,\ 1.5)$, $(+1,\ 0.3)$, $(-1,\ -2.0)$ and $(-1,\ 0.7)$, labelling each case as correct-with-margin, margin violation or misclassified.

Q3. [4 Marks]
For $\mathbf{x} = (2, 1)$ and $\mathbf{z} = (1, 3)$ compute the polynomial kernel $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^\top\mathbf{z} + 1)^2$ and the RBF kernel $\exp(-\gamma\|\mathbf{x} - \mathbf{z}\|^2)$ with $\gamma = 0.5$.

Q4. [4 Marks]
A neural network has $M = 4$ hidden units with tanh activation. Compute the number of weight vectors that give exactly the same input–output mapping, and explain the two symmetries involved.

Q5. [4 Marks]
Compute the entropy of a node containing (a) 10 positive and 10 negative examples, (b) 8 positive and 2 negative, and (c) 10 positive and 0 negative. Comment on the three values.

Q6. [4 Marks]
In bagging with a bootstrap sample of size $m = 500$, compute the probability that a given example is *not* selected, and state the limiting out-of-bag fraction as $m \to \infty$.

SECTION B — 8 Mark Questions (Attempt all four)

Q7. [8 Marks]
Derive the *maximum-margin* formulation of a hard-margin SVM, showing why maximising $2/\|\mathbf{w}\|$ is equivalent to minimising $\frac{1}{2}\|\mathbf{w}\|^2$. For the two points $\mathbf{x}_1 = (2, 2)$ with $y_1 = +1$ and $\mathbf{x}_2 = (0, 0)$ with $y_2 = -1$, determine $\mathbf{w}$, b and the margin.

Q8. [8 Marks]
Explain the *soft-margin* SVM with slack variables. Show that it is equivalent to regularized empirical risk minimization with the hinge loss, and explain the effect of C on bias and variance.

Q9. [8 Marks]
A node contains 10 positive and 10 negative examples. A candidate attribute splits it into two children of 10 examples each: $(8+, 2-)$ and $(2+, 8-)$.

(a) Compute the entropy of the parent and of each child.

(b) Compute the information gain of the split.

(c) Compute the Gini index of the parent and the Gini reduction, and compare the two criteria.

Q10. [8 Marks]
Explain *bagging* and derive the variance of the average of T models each of variance σ^2 with pairwise correlation ρ. Evaluate it for $\sigma^2 = 1$ with $(\rho, T) = (0, 20)$ and $(0.4, 20)$, and explain what this implies for random forests.

SECTION C — 10 Mark Questions (Attempt all three)

Q11. [10 Marks]

(a) Derive the *dual* of the hard-margin SVM from the Lagrangian, and state the KKT complementary-slackness condition. (6)

(b) Explain what $\alpha_i > 0$ signifies, and state what happens to the classifier if a point with $\alpha_i = 0$ is removed from the training set. (4)

Q12. [10 Marks]
A single sigmoid neuron has inputs $\mathbf{x} = (1, 2)$, weights $\mathbf{w} = (0.3, 0.5)$, bias $b = 0.1$ and target $t = 1$, with squared-error loss $E = \frac{1}{2}(t - o)^2$.

(a) Perform the forward pass: compute the activation a, the output o and the error E. (4)

(b) Perform the backward pass: compute $\partial E/\partial o$, $\partial o/\partial a$, the error signal δ, and $\partial E/\partial w_1$ and $\partial E/\partial w_2$. (4)

(c) Update both weights using $\eta = 0.5$. (2)

[$\sigma(1.4) = 0.8022$.]

Q13. [10 Marks]

Explain *overfitting* in decision trees and sketch the accuracy-versus-tree-size curve. Describe pre-pruning and post-pruning, explain reduced-error pruning and rule post-pruning, and state the Minimum Description Length principle as applied to trees.

SECTION D — 12 Mark Questions (Attempt all three)

Q14. **[12 Marks]**

(a) Explain the *kernel trick* and state Mercer's condition. (4)

(b) For $\mathbf{x} = (1, 2)$ and $\mathbf{z} = (3, 4)$, compute $(\mathbf{x}^\top \mathbf{z} + 1)^2$ directly, and then verify it by explicitly computing $\phi(\mathbf{x})^\top \phi(\mathbf{z})$ with $\phi(\mathbf{x}) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$. (5)

(c) Name four valid kernels and state one situation in which each is preferred. (3)

Q15. **[12 Marks]**

(a) Derive the backpropagation equations, obtaining $\delta_j = h'(a_j) \sum_k w_{kj}\delta_k$ and $\partial E/\partial w_{ji} = \delta_j z_i$. (6)

(b) State the computational cost of backpropagation and of numerical differentiation for the full gradient, in terms of the number of weights W, and justify the difference. (4)

(c) Explain the purpose of gradient checking. (2)

Q16. **[12 Marks]**

(a) Explain the ID3 algorithm, defining entropy, information gain and gain ratio, and state the bias of information gain. (5)

(b) Explain how random forests are obtained from bagging, defining m_{try} and giving its default values for classification and regression when $p = 16$ and $p = 100$. (4)

(c) Explain out-of-bag error and the two measures of variable importance. (3)

4 Unit 4 — Unsupervised Learning

Unit 4: Unsupervised Learning

A. What is unsupervised learning; clustering methods (distance-based, probabilistic, hierarchical). **B.** k-means and initialization; applications; GMM and EM. **C.** Principal Component Analysis; properties, limitations and applications.

SECTION A — 4 Mark Questions (Attempt all six)

Q1. **[4 Marks]**
For the points $(1, 2)$ and $(4, 6)$ compute the Euclidean, Manhattan and Chebyshev distances, writing the Minkowski formula from which all three follow.

Q2. **[4 Marks]**
A point has mean intra-cluster distance $a(i) = 3$ and mean nearest-other-cluster distance $b(i) = 5$. Compute its silhouette coefficient and interpret it. Repeat for $a(i) = 5$, $b(i) = 3$.

Q3. **[4 Marks]**
A cluster contains the points $(2, 4)$, $(4, 6)$ and $(6, 8)$. Compute the centroid and the cluster's contribution to the SSE.

Q4. **[4 Marks]**
Write the four standard linkage formulas used in agglomerative clustering (single, complete, average and Ward), and state which of them tends to produce chained, elongated clusters.

Q5. **[4 Marks]**
State the responsibility formula $\gamma(z_{ik})$ of a Gaussian mixture model, explain what it must sum to over k, and state the limiting condition under which EM reduces to k-means.

Q6. **[4 Marks]**
The covariance matrix of a two-feature dataset is $\mathbf{S} = \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}$. Compute both eigenvalues and the percentage of variance explained by the first principal component.

SECTION B — 8 Mark Questions (Attempt all four)

Q7. **[8 Marks]**
Five points lie on a line at $P_1 = 2$, $P_2 = 5$, $P_3 = 9$, $P_4 = 10$, $P_5 = 15$.

(a) Construct the 5×5 distance matrix.

(b) Perform agglomerative clustering with *single* linkage, listing every merge and its distance.

(c) Repeat with *complete* linkage and explain why the two dendrograms differ.

Q8. **[8 Marks]**
Derive the two update steps of k-means from the distortion $J = \sum_i \sum_k r_{ik} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2$, showing that the optimal prototype is the cluster mean. State the convergence guarantee and the time complexity.

Q9. **[8 Marks]**
For the points $A(2, 2)$, $B(3, 3)$, $C(8, 8)$, $D(9, 9)$, $E(2, 4)$ with initial centroids $\boldsymbol{\mu}_1 = (2, 2)$ and $\boldsymbol{\mu}_2 = (8, 8)$:

(a) Assign every point to its nearest centroid, showing the distances.

(b) Compute the updated centroids.

(c) Compute the distortion J after the update.

Q10. **[8 Marks]**
Explain the *k-means++* initialization procedure. With the first centre chosen at $A(2, 2)$ and the remaining points $B(3, 3)$, $C(8, 8)$, $D(9, 9)$, $E(2, 4)$, compute $D(\mathbf{x})^2$ for each point and the probability that each is chosen as the second centre.

SECTION C — 10 Mark Questions (Attempt all three)

Q11. **[10 Marks]**

(a) Explain the Gaussian mixture model and write the E-step and M-step equations of the EM algorithm. (5)

(b) For the 1-D data $X = \{2, 3, 7, 8\}$ with initial $\mu_1 = 3$, $\mu_2 = 7$, $\sigma_1^2 = \sigma_2^2 = 4$ and $\pi_1 = \pi_2 = 0.5$, compute the responsibility γ_{i1} of every point. (5)

[$\sqrt{2\pi(4)} = 5.0133$.]

Q12. **[10 Marks]**
Derive the first principal component using a Lagrange multiplier, showing that the optimal direction satisfies $\mathbf{S}\mathbf{u} = \lambda\mathbf{u}$ and that the variance captured equals λ. Explain how further components are obtained and why they are orthogonal.

Q13. **[10 Marks]**

(a) State four properties and four limitations of PCA. (5)

(b) Explain the relationship between PCA and the SVD, deriving $\lambda_k = \sigma_k^2/(m-1)$. (3)

(c) A dataset of $m = 10$ samples has singular values $\sigma_1 = 6$ and $\sigma_2 = 4$. Compute λ_1 and λ_2. (2)

SECTION D — 12 Mark Questions (Attempt all three)

Q14. **[12 Marks]**

Consider the dataset $(2, 1),\ (3, 3),\ (4, 2),\ (5, 4),\ (6, 5)$.

(a) Compute the mean vector and centre the data. (3)

(b) Compute the 2×2 covariance matrix. (3)

(c) Compute both eigenvalues, verifying them against the trace, and give the percentage of variance explained by each component. (4)

(d) Compute the unit eigenvector of the larger eigenvalue and verify that the two eigenvectors are orthogonal. (2)

Q15. **[12 Marks]**

(a) Compare k-means with the Gaussian mixture model on any six points. (5)

(b) State five limitations of k-means and give one remedy for each. (5)

(c) Explain why k-means fails on two concentric rings, with a sketch. (2)

Q16. **[12 Marks]**

(a) Explain *vector quantization* and how k-means builds the codebook. (3)

(b) A 512×512 image at 24-bit colour is quantized to $K = 256$ colours. Compute the original size, the codebook size, the index storage and the compression ratio. (5)

(c) Explain two further applications of clustering: cluster analysis and feature learning. (4)

5 Unit 5 — Parameter Estimation, Model Evaluation and Ensembles

Unit 5: Parameter Estimation, Model Evaluation and Ensemble Methods
A. Point estimation, MLE, unbiased estimation, confidence intervals. **B.** Model validation methods and the ROC curve. **C.** Ensemble theory, size, voting/averaging, bagging, boosting, stacking.

SECTION A — 4 Mark Questions (Attempt all six)

Q1. **[4 Marks]**
In 25 Bernoulli trials there are 18 successes. Compute the maximum-likelihood estimate of p and state the general formula, deriving it briefly.

Q2. **[4 Marks]**
For the sample $3,\ 5,\ 5,\ 7,\ 10$, compute the sample mean, the MLE of the variance and the unbiased estimate of the variance. Explain in one line why the two differ.

Q3. **[4 Marks]**
A sample of $n = 36$ has $\bar{x} = 70$ and known $\sigma = 12$. Construct the 95% confidence interval for μ. [$z_{0.025} = 1.96$.]

Q4. **[4 Marks]**
A classifier gives $TP = 45$, $FN = 15$, $FP = 25$, $TN = 415$. Compute the accuracy, precision, recall and F_1 score.

Q5. **[4 Marks]**
Five-fold cross-validation gives the accuracies $0.78,\ 0.82,\ 0.80,\ 0.86,\ 0.84$. Compute the mean and the sample standard deviation, and state how the result should be reported.

Q6. **[4 Marks]**
In AdaBoost a weak learner has weighted error $\epsilon_t = 0.25$. Compute α_t, and state the value of α_t when $\epsilon_t = 0.5$ and what it means. [$\ln 3 = 1.0986$.]

SECTION B — 8 Mark Questions (Attempt all four)

Q7. **[8 Marks]**
Explain the method of *maximum likelihood*. Derive the MLE of the Bernoulli parameter p from first principles, verifying that the stationary point is a maximum. Then state (without full derivation) the MLE for the Poisson and exponential distributions, and evaluate them for the data $2, 4, 3, 6, 5$ and $1, 2, 3, 4$ respectively.

Q8. **[8 Marks]**
Explain the meaning of a 95% confidence interval, stating clearly what is random and what is fixed. For the sample $8,\ 10,\ 12,\ 9,\ 11$, compute $\bar{x}$, s, the standard error and the 95% t-interval for μ. [$t_{0.025,4} = 2.776$.]

Q9. **[8 Marks]**
Explain *bootstrap* validation. Compute the out-of-bag fraction for $m = 50$, $m = 200$ and $m \to \infty$. Given $\mathrm{Err}_{\mathrm{train}} = 0.04$ and $\mathrm{Err}_{\mathrm{OOB}} = 0.28$, compute the .632 estimator and explain why the blend is used.

Q10. **[8 Marks]**
State Condorcet's jury theorem and its two conditions. Compute the majority-vote error of an ensemble of $T = 3$ independent classifiers each with error $p = 0.25$, showing both terms of the sum, and repeat for $T = 5$ with $p = 0.2$.

SECTION C — 10 Mark Questions (Attempt all three)

Q11. **[10 Marks]**
Eight test examples, sorted by predicted score, have the true labels shown:

Score	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
True	+	+	−	+	−	+	−	−

(a) Construct the ROC table giving TP, FP, TPR and FPR at every threshold. (5)

(b) Sketch the ROC curve on labelled axes. (2)

(c) Compute the AUC by the trapezoidal method *and* verify it by the Mann–Whitney pair-counting method. (3)

Q12. **[10 Marks]**
Compare the validation methods — holdout, k-fold cross-validation, leave-one-out, random subsampling, teach-and-test and bootstrap — on the number of models trained, bias, variance and best use case. Explain why the test set must be touched only once, and what nested cross-validation is used for.

Q13. **[10 Marks]**

(a) Explain the AdaBoost algorithm, writing the formulas for ϵ_t, α_t and the weight update. (5)

(b) Four training points carry uniform weights 0.25. A weak learner misclassifies exactly one of them. Compute ϵ_1, α_1, the unnormalised weights, the normalising constant Z_1 and the updated weights. (5)

[$e^{0.5493} = 1.7321$, $e^{-0.5493} = 0.5774$.]

SECTION D — 12 Mark Questions (Attempt all three)

Q14. **[12 Marks]**

(a) Define bias, variance and mean squared error of an estimator, and state the relation between them. (3)

(b) Prove that $\mathbb{E}\left[\frac{1}{n}\sum(X_i - \bar{X})^2\right] = \frac{n-1}{n}\sigma^2$, and hence explain the $n-1$ divisor. (5)

(c) Two independent samples give $n_1 = 8$, $\bar{x}_1 = 85$, $s_1 = 4$ and $n_2 = 10$, $\bar{x}_2 = 80$, $s_2 = 3$. Compute the pooled variance, the standard error and the 95% confidence interval for the difference of means, and state your conclusion. (4)

[$t_{0.025,16} = 2.120$.]

Q15. **[12 Marks]**

(a) Explain the confusion matrix and define accuracy, precision, recall, specificity and F_1. (4)

(b) For $TP = 45$, $FN = 15$, $FP = 25$, $TN = 415$, compute all five measures and also the accuracy of a trivial "always negative" classifier. Comment on the comparison. (5)

(c) Explain what AUC $= 0.5$ signifies and when a precision–recall curve should be preferred to an ROC curve. (3)

Q16. **[12 Marks]**

(a) Explain ensemble theory and derive the variance of an average of T correlated models. (4)

(b) Three classifiers output $P(\text{class } 1) = 0.40,\ 0.45,\ 0.85$. Determine the prediction under *hard* voting and under *soft* voting, and explain why they differ. (4)

(c) Three regressors predict $20,\ 24,\ 30$ with weights $0.5,\ 0.3,\ 0.2$. Compute the weighted average and the simple average. Then compare bagging with boosting on any four points. (4)

Appendix — Answer Key for Numerical Questions

For instructor use. Final answers only — students must show complete working to earn full marks.

Unit 1

- **Q3.** min–max $= \frac{45-20}{120-20} = \mathbf{0.25}$; $z = \frac{45-60}{25} = \mathbf{-0.60}$.
- **Q4.** errors $(-1, 1, -1, 1)$; MSE $= \mathbf{1.0}$, RMSE $= \mathbf{1.0}$, MAE $= \mathbf{1.0}$.
- **Q5.** support $= \frac{240}{2000} = \mathbf{0.12}$; confidence $= \frac{240}{600} = \mathbf{0.40}$; lift $= \frac{0.40}{0.20} = \mathbf{2.0}$ (positive association).
- **Q6.** $2^8 = \mathbf{256}$; $2^{15} = \mathbf{32768}$.
- **Q9.** Equal-frequency bins $\{5, 10, 11\}, \{13, 15, 35\}, \{50, 55, 72\}$.
  Means: **8.67**, **21**, **59**.
  Boundaries: $(5, 11, 11)$, $(13, 13, 35)$, $(50, 50, 72)$.
  Equal width $= \frac{72-5}{3} = \mathbf{22.33}$: bin1 $\{5, 10, 11, 13, 15\}$, bin2 $\{35\}$, bin3 $\{50, 55, 72\}$.
- **Q10.** $\sum a_i b_i - n\bar{A}\bar{B} = 66 - 60 = 6$; Cov $= \frac{6}{5} = \mathbf{1.20}$; $r = \frac{6}{5(1.4142)(1.0954)} = \mathbf{0.7746}$ (strong positive).
- **Q12.** Expected: 40, 60, 40, 60. Contributions: 10.0, 6.6667, 10.0, 6.6667; $\chi^2 = \mathbf{33.3333} > 3.841 \Rightarrow$ reject independence (attributes correlated).
- **Q13.** mean $= \frac{22+28+34+36}{4} = \mathbf{30}$ (thousands).
- **Q15(c).** accuracy $= \frac{980}{1000} = \mathbf{98\%}$ while detecting zero fraud.

Unit 2

- **Q1.** MSE $= 4 + 3^2 = \mathbf{13}$; unbiased case $= 9 + 0 = \mathbf{9}$ (prefer the second).
- **Q2.** 10^n: **10**, **1000**, 10^6 — curse of dimensionality.
- **Q3.** $\theta_1 = \frac{15}{10} = \mathbf{1.5}$; $\theta_0 = 5.8 - 1.5(3) = \mathbf{1.3}$.
- **Q4.** min–max $= \frac{1500-900}{900} = \mathbf{0.6667}$; mean-norm $= \frac{100}{900} = \mathbf{0.1111}$; $z = \frac{100}{316.2278} = \mathbf{0.3162}$.
- **Q5.** $\sigma(-1) = \mathbf{0.2689}$, $\sigma(0) = \mathbf{0.5}$, $\sigma(1) = \mathbf{0.7311}$; odds $= \frac{0.8}{0.2} = \mathbf{4}$; logit $= \ln 4 = \mathbf{1.3863}$.
- **Q6.** boundary $x = \frac{3}{0.06} = \mathbf{50}$; OR $= \mathbf{1.0618}$.
- **Q7.** $\hat{y} = (2.8, 4.3, 5.8, 7.3, 8.8)$; $e = (0.2, -0.3, 0.2, -0.3, 0.2)$, $\sum e_i = 0$ ✓
  SSE $= \mathbf{0.30}$, SST $= \mathbf{22.80}$, $R^2 = \mathbf{0.9868}$, MSE $= \mathbf{0.06}$.
- **Q8.** det $= 50$; $\boldsymbol{\theta} = \frac{1}{50}\begin{pmatrix} 55 & -15 \\ -15 & 5 \end{pmatrix}\begin{pmatrix} 29 \\ 102 \end{pmatrix} = (\mathbf{1.3},\ \mathbf{1.5})^\top$ ✓
- **Q9.** $\theta = \frac{30}{15+\lambda}$: $\lambda = 0 \to \mathbf{2.0}$, $\lambda = 3 \to \mathbf{1.6667}$, $\lambda = 15 \to \mathbf{1.0}$.
- **Q10.** odds(treated) $= \frac{45}{15} = 3$; odds(control) $= \frac{20}{60} = 0.3333$; OR $= \mathbf{9.0}$ ($= \frac{ad}{bc} = \frac{45\times60}{15\times20}$).
- **Q11(b).** $J_0 = \mathbf{19.10}$. Iter 1: $g_0 = -5.8$, $g_1 = -20.4 \Rightarrow \theta = (\mathbf{0.29},\ \mathbf{1.02})$, $J = \mathbf{3.2617}$.
  Iter 2: $g_0 = -2.5$, $g_1 = -8.3 \Rightarrow \theta = (\mathbf{0.4125},\ \mathbf{1.4355})$, $J = \mathbf{0.6184}$.
- **Q12(b).** losses 0.2231, 0.5108, 0.3567, 0.1054; average $= \mathbf{0.2990}$.
- **Q13.** $\hat{\mu}_1 = \mathbf{2.5}$, $\hat{\mu}_2 = \mathbf{7.5}$, $s_p^2 = \mathbf{1.6667}$; threshold $= \mathbf{5.0}$; $x = 4.5 \to$ class **1**, $x = 5.5 \to$ class **2**; $D^2 = \frac{25}{1.6667} = \mathbf{15.0}$, $D = 3.873$ (well separated).
- **Q16(b).** $x = 30:\ z = -1.2,\ p = \mathbf{0.2315}$, odds 0.3012; $x = 50:\ z = 0,\ p = \mathbf{0.5}$; $x = 70:\ z = 1.2,\ p = \mathbf{0.7685}$, odds 3.3201.

Unit 3

- **Q1.** $\|\mathbf{w}\| = 5$, margin $= \mathbf{0.4}$; $\|\mathbf{w}\| = 1$, margin $= \mathbf{2.0}$.
- **Q2.** $L = \mathbf{0}$, $\mathbf{0.7}$, $\mathbf{0}$, $\mathbf{1.7}$ (correct-with-margin, margin violation, correct-with-margin, misclassified).
- **Q3.** $\mathbf{x}^\top\mathbf{z} = 5$, $K_{\text{poly}} = 6^2 = \mathbf{36}$; $\|\mathbf{x} - \mathbf{z}\|^2 = 5$, $K_{\text{RBF}} = e^{-2.5} = \mathbf{0.0821}$.
- **Q4.** $M!\,2^M = 24 \times 16 = \mathbf{384}$.

- **Q5.** $H(10,10) = \mathbf{1.0}$; $H(8,2) = \mathbf{0.7219}$; $H(10,0) = \mathbf{0}$.
- **Q6.** $(1-1/500)^{500} = \mathbf{0.3675}$; limit $1/e = \mathbf{0.3679} \Rightarrow 36.8\%$.
- **Q7.** $b = \mathbf{-1}$, $\mathbf{w} = \mathbf{(0.5,0.5)}$, $\|\mathbf{w}\| = 0.7071$, margin $= \mathbf{2.8284} = 2\sqrt{2}$.
- **Q9.** $E(\text{parent}) = 1.0$; each child $= 0.7219$; remainder $= 0.7219$; Gain $= \mathbf{0.2781}$.
  Gini parent $= \mathbf{0.5}$, after split $= \mathbf{0.32}$, reduction $= \mathbf{0.18}$.
- **Q10.** Var $= \rho\sigma^2 + \frac{1-\rho}{T}\sigma^2$: $(0,20) \to \mathbf{0.05}$; $(0.4,20) \to \mathbf{0.43}$ — correlation limits the gain, hence feature subsampling.
- **Q12.** $a = 0.3(1) + 0.5(2) + 0.1 = \mathbf{1.4}$; $o = \mathbf{0.8022}$; $E = \mathbf{0.019566}$.
  $\partial E/\partial o = -0.1978$, $o(1-o) = 0.1587$, $\delta = \mathbf{-0.03139}$.
  $\partial E/\partial w_1 = \mathbf{-0.03139}$, $\partial E/\partial w_2 = \mathbf{-0.06278}$;
  $w_1^+ = \mathbf{0.3157}$, $w_2^+ = \mathbf{0.5314}$.
- **Q14(b).** $\mathbf{x}^\top\mathbf{z} = 11$, $K = 12^2 = \mathbf{144}$;
  $\phi(\mathbf{x}) = (1, 1.414, 2.828, 1, 4, 2.828)$, $\phi(\mathbf{z}) = (1, 4.243, 5.657, 9, 16, 16.971)$, dot $= \mathbf{144}$ ✓
- **Q16(b).** m_{try}: $p = 16 \to \sqrt{16} = \mathbf{4}$ (class.), $16/3 = \mathbf{5}$ (regr.); $p = 100 \to \mathbf{10}$ and $\mathbf{33}$.

Unit 4

- **Q1.** Euclidean $= \sqrt{9+16} = \mathbf{5}$; Manhattan $= \mathbf{7}$; Chebyshev $= \mathbf{4}$.
- **Q2.** $s = \frac{5-3}{5} = \mathbf{0.4}$ (well placed); $s = \frac{3-5}{5} = \mathbf{-0.4}$ (wrong cluster).
- **Q3.** centroid $= \mathbf{(4,6)}$; SSE $= 8 + 0 + 8 = \mathbf{16}$.
- **Q6.** $(4-\lambda)^2 = 4 \Rightarrow \lambda = \mathbf{6},\ \mathbf{2}$; PC1 $= \frac{6}{8} = \mathbf{75\%}$.
- **Q7.** Single linkage merges: $\{P_3, P_4\}$ at 1; $\{P_1, P_2\}$ at 3; join at 4; P_5 at 5.
  Complete linkage: $\{P_3, P_4\}$ at 1; $\{P_1, P_2\}$ at 3; P_5+$\{P_3P_4\}$ at $\mathbf{6}$; final at $\mathbf{13}$.
- **Q9.** Clusters $\{A, B, E\}$ and $\{C, D\}$; new centroids $\mathbf{(2.3333,3.0)}$ and $\mathbf{(8.5,8.5)}$; $J = \mathbf{3.6667}$.
- **Q10.** D^2: A=0, B=2, C=72, D=98, E=4; total $= \mathbf{176}$.
  $P = 0$, $\mathbf{0.0114}$, $\mathbf{0.4091}$, $\mathbf{0.5568}$, $\mathbf{0.0227}$.
- **Q11(b).** γ_1: x=2 $\to \mathbf{0.9526}$; x=3 $\to \mathbf{0.8808}$; x=7 $\to \mathbf{0.1192}$; x=8 $\to \mathbf{0.0474}$.
- **Q13(c).** $\lambda_1 = \frac{36}{9} = \mathbf{4.0}$; $\lambda_2 = \frac{16}{9} = \mathbf{1.7778}$.
- **Q14.** mean $= \mathbf{(4,3)}$; centred $(-2,-2), (-1,0), (0,-1), (1,1), (2,2)$;
  $\mathbf{S} = \begin{pmatrix} 2.5 & 2.25 \\ 2.25 & 2.5 \end{pmatrix}$; $\lambda_1 = \mathbf{4.75}$, $\lambda_2 = \mathbf{0.25}$ (sum $= 5 =$ trace ✓);
  variance explained $\mathbf{95\%}$ and $\mathbf{5\%}$; $\mathbf{u}_1 = \mathbf{(0.7071,0.7071)}$, $\mathbf{u}_2 = (0.7071, -0.7071)$, $\mathbf{u}_1^\top\mathbf{u}_2 = 0$ ✓
- **Q16(b).** original $= 512^2 \times 24 = \mathbf{6{,}291{,}456}$ bits (768 KB); codebook $= 256 \times 24 = 6144$; indices $= 512^2 \times 8 = 2{,}097{,}152$; total $= 2{,}103{,}296$ bits (256.8 KB); ratio $= \mathbf{2.99{:}1}$.

Unit 5

- **Q1.** $\hat{p} = \frac{18}{25} = \mathbf{0.72}$.
- **Q2.** $\bar{x} = \mathbf{6}$; $\sum(x-\bar{x})^2 = 28$; MLE $= \frac{28}{5} = \mathbf{5.6}$; unbiased $= \frac{28}{4} = \mathbf{7.0}$.
- **Q3.** SE $= \frac{12}{6} = 2$; MOE $= 3.92$; CI $= \mathbf{(66.08,\ 73.92)}$.
- **Q4.** Accuracy $= \frac{460}{500} = \mathbf{0.92}$; Precision $= \frac{45}{70} = \mathbf{0.6429}$; Recall $= \frac{45}{60} = \mathbf{0.75}$; $F_1 = \mathbf{0.6923}$.
- **Q5.** mean $= \mathbf{0.82}$; $\sum(a-\bar{a})^2 = 0.004$; $s = \mathbf{0.0316}$; report 0.820 ± 0.032.
- **Q6.** $\alpha = \frac{1}{2}\ln 3 = \mathbf{0.5493}$; at $\epsilon = 0.5$, $\alpha = \mathbf{0}$ (the learner gets no vote).
- **Q7.** Poisson $\hat{\lambda} = \frac{20}{5} = \mathbf{4}$; exponential $\hat{\lambda} = \frac{4}{10} = \mathbf{0.4}$.
- **Q8.** $\bar{x} = \mathbf{10}$; $\sum(y-\bar{y})^2 = 10$; $s^2 = \mathbf{2.5}$, $s = 1.5811$; SE $= \mathbf{0.7071}$; MOE $= 1.9629$; CI $= \mathbf{(8.0371,\ 11.9629)}$.
- **Q9.** OOB: m=50 $\to \mathbf{0.3642}$, m=200 $\to \mathbf{0.3670}$, limit $\mathbf{0.3679}$;
  $\text{Err}_{.632} = 0.368(0.04) + 0.632(0.28) = \mathbf{0.1917}$.
- **Q10.** T=3, p=0.25: $3(0.0625)(0.75) + 0.015625 = 0.1406 + 0.0156 = \mathbf{0.1562}$.
  T=5, p=0.2: $\mathbf{0.0579}$.

- **Q11.** $P = 4$, $N = 4$. The (FPR, TPR) points reached are $(0, .25),\ (0, .5),\ (.25, .5),\ (.25, .75),\ (.5, .75),\ (.5, 1),\ (.75, 1),\ (1, 1)$.
AUC = **0.8125** by area; Mann–Whitney = $\frac{13}{16}$ = **0.8125** ✓
- **Q13(b).** ϵ_1 = **0.25**; α_1 = **0.5493**; wrong = 0.25(1.7321) = 0.4330, correct = 0.25(0.5774) = 0.1443; Z_1 = **0.8660**;
normalised: wrong = **0.5**, each correct = **0.1667** (sum = 1 ✓).
- **Q14(c).** $s_p^2 = \frac{7(16)+9(9)}{16}$ = **12.0625**, $s_p = 3.4731$; SE = **1.6474**; MOE = 3.4926; CI = **(1.5074, 8.4926)** — excludes 0, so the difference is significant.
- **Q15(b).** Specificity = $\frac{415}{440}$ = **0.9432**; trivial classifier accuracy = $\frac{440}{500}$ = **0.88** (close to 0.92, so accuracy alone is misleading).
- **Q16(b).** Hard vote 2–1 for class **0**; soft vote mean = $\frac{0.40+0.45+0.85}{3}$ = **0.5667** $> 0.5 \Rightarrow$ class **1**.
- **Q16(c).** weighted = $0.5(20) + 0.3(24) + 0.2(30)$ = **23.2**; simple = **24.6667**.

End of Assignment Question Bank — Machine Learning (CSE473)
Post your doubts: gcjana.in/.../machine-learning/#Post-doubts