

## Unit 1

1. What is machine learning? Why do we need it? Explain the application areas of the machine learning in real world.
2. Explain different types of machine learning methods: Supervised learning, Unsupervised learning, Reinforcement learning and active learning with examples.
3. Explain the machine learning workflow in details with a neat and clean diagram.
4. Explain what are the problems in machine learning. Why ethical consideration must be done for the machine learning. Explain with examples
5. What are the different types of data quality issues in machine learning. Explain with examples.
6. Why do we need machine learning? Differentiate between noise and error.

## Unit 2

7. What is supervised learning. Explain the workflow of supervised learning with examples.
8. Differentiate between regression and classification task with examples.
9. Compare between linear, multiple and polynomial regression with examples
10. For the data in the given table, find the equation of linear regression.

Experience (in years)	3	5	4	12	10	7	8
Salary (in \$)	3000	5000	4000	11000	8900	6700	7100

11. For the given data, find the equation of the best fit lines using Linear Regression.

Years of Experience (X1)	2	3	5	7	10
Education Level (X2)	12	14	16	18	20
Salary (Y)	45000	65000	67000	75000	90000

12. Compare between linear and logistic regression. Explain with examples
13. How do you find the best fit line in linear regression. Explain
14. Using the logistic regression model and a probability cutoff of 0.75, classify the following customers as "Owner" or "Nonowner." If  $p \geq 0.75$ , classify as "Owner"; otherwise, as "Nonowner." Assume  $\beta_0 = -25.9382$ ,  $\beta_1 = 0.1109(\text{Income})$ ,  $\beta_2 = 0.9638(\text{Lot Size})$

Customer	Income	Lot Size	Ownership
1	60	18.4	Owner
2	64.8	21.6	Owner

3	84	17.6	Nonowner
4	59.4	20	Nonowner
5	108	17.6	Owner
6	75	19.6	Nonowner

15. What is learning rate? Why do we need learning rate?

16. For the data in the given table, find the class of the new data using KNN algorithm.

Name	Age	Income	No. of Credit	Class
Karina	20	5000	3	Yes
Shriya	21	4000	2	No
Riya	22	3000	1	Yes
Saugat	23	2500	3	No
Dil	19	3500	2	Yes
Priya	24	1700	1	Yes
Junet	22	2200	2	No
Manjil	25	3100	3	?

17. What is regularization? Why do we need it?

18. Explain the different types of regularization techniques with examples.

19. Write algorithm for logistic regression.

20. What is a decision tree? Explain the steps involved in creating a decision tree. Also state its limitations.

21. Explain the working principle of Decision tree and also explain how the feature selection is done in decision tree.

22. What is instance based learning? Explain the KNN algorithm with an example.

23. What is entropy and gini index in the decision tree?

24. Create Decision tree using ID3 algorithm.

Age	Job	House	Credit	Loan Approved
Young	FALSE	No	Fair	No
Young	FALSE	No	Good	No
Young	TRUE	No	Good	Yes

Young	TRUE	Yes	Fair	Yes
Young	FALSE	No	Fair	No
Middle	FALSE	No	Fair	No
Middle	FALSE	No	Good	No
Middle	TRUE	Yes	Good	Yes
Middle	FALSE	Yes	Excellent	Yes
Middle	FALSE	Yes	Excellent	Yes
Old	FALSE	Yes	Excellent	Yes
Old	FALSE	Yes	Good	Yes
Old	TRUE	No	Good	Yes
Old	TRUE	No	Excellent	Yes
Old	FALSE	No	Fair	No

25. What is random forest classifier? Explain how random forest overcomes the limitation of decision tree.
26. Explain the ensemble method in random forest classifier and explain how it works.
27. How can you find the optimal solution of k in KNN algorithm?
28. What is SVM? Explain the steps to implement a linear SVM classifier.
29. Explain the different types kernel functions used in SVM classifier.
30. What are support vectors in SVM? How they are used to define the margins between data points. Also explain how maximum separation is calculated? Compare between functional and geometric margin.
31. Explain how SVM works? Find the line of separation for the following data using SVM.

X1	4	4	7	7	1	1	-1	-1
X2	2	-2	1	-1	1	-1	1	-1
Label	0	0	0	0	1	1	1	1

32. Create line of separation using Linear SVM for the following dataset

Point	x1	x2	y
A	2	2	-1
B	4	5	1

C	7	4	1
D	3	3	-1
E	6	6	1
F	8	5	-1

### Unit 3

33. What is unsupervised learning? Explain the application areas of clustering in real world.
34. What is K in k-means clustering? Explain the steps involved in K-means clustering.
35. Explain steps used in K means clustering. Apply K(=2)-Means algorithm over the data (185, 72), (170, 56), (168, 60), (179,68), (182,72), (188,77) up to two iterations and show the clusters. Initially choose first two objects as initial centroids.
36. For the following dataset create clusters using k means clustering. Assume k=3.

Point	A1	A2	A3	A4	A5	A6	A7	A8
x <sub>1</sub>	2	2	8	5	7	6	1	4
x <sub>2</sub>	10	5	4	8	5	4	2	9

37. Define the steps involved in calculating agglomerative hierarchical clustering. Also explain the advantages and disadvantages of hierarchical clustering.
38. Create dendrogram from the given distance matrix. Use both single and complete linkage technique.

	A	B	C	D	E	F
A	0	4	6	5	8	10
B	4	0	6	7	9	11
C	6	6	0	3	4	6
D	5	7	3	0	5	7
E	8	9	4	5	0	3
F	10	11	6	7	3	0

39. Explain the use of dimensionality reduction in machine learning? Explain how PCA works.
40. What is density based clustering. Explain its uses.
41. Compare between PCA and LDA with examples.
42. For the given data in the table, convert the two dimensional data into a single dimension using PCA algorithm.

Feature	Example 1	Example 2	Example 3	Example 4
X1	4	8	13	7
X2	11	4	5	14

Also, explain why PCA is needed

#### Unit 4

43. What is ANN? Explain its components with a diagram.
44. What is perceptron. Compare single layer and multi layer perceptron with examples.
45. What is a feed forward and feed backward networks with examples.
46. What is convolution? Explain different types of activation function in ANN.
47. Design a perceptron for OR gate and find the appropriate weights and bias for this model.
48. Design a perceptron for XOR gate and find the appropriate weights and bias for this model.
49. What is gradient descent? Explain the steps involved in gradient descent.
50. How does gradient descent optimizes the algorithm.
51. Define bias and variance with examples. Explain the bias-variance tradeoff.
52. What is overfitting and underfitting? Explain with example.
53. What is CNN? Explain its components and its architecture.
54. What does dropout and batch normalization do in CNN?
55. How CNN can be used in Computer Vision and Image processing tasks.
56. Consider a grayscale input image of size  $7 \times 7$  pixels. The input image converted into matrix is

$$\begin{bmatrix} 1 & 2 & 0 & 1 & 3 & 1 & 2 \\ 4 & 1 & 3 & 0 & 1 & 2 & 1 \\ 1 & 0 & 2 & 3 & 1 & 0 & 2 \\ 2 & 1 & 0 & 1 & 2 & 3 & 1 \\ 0 & 3 & 1 & 2 & 0 & 1 & 2 \\ 1 & 2 & 3 & 1 & 0 & 2 & 1 \\ 2 & 1 & 0 & 2 & 3 & 1 & 0 \end{bmatrix}$$

And the size of the kernel is  $3 \times 3$  which is

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Consider a stride of 2 and padding 1, show the step by step calculation for extracting features from the data.

57. What is RNN? Explain its component along with its architecture.
58. What is LSTM? Explain its components and describe how they work?

- 59. What is GRU? Explain its components and describe how they work?
- 60. What is a time series data? Explain the application areas of time series forecasting.

### **Unit 5**

- 61. Why do we need to evaluate the model in Machine Learning?
- 62. Explain training data, validation data and testing data with examples.
- 63. What is confusion matrix? Explain TP, TN, FP and FN with examples.
- 64. Differentiate between specificity and sensitivity? Also explain False Positive Rate.
- 65. Explain the term accuracy, precision, recall and f1-score with examples.
- 66. What is a PR curve (Precision-Recall curve). When can this be used?
- 67. Explain ROC and AUC curve with example and how they can be used to evaluate model.
- 68. How can you evaluate a regression model. Explain the uses of MSE, MAE, MAPE, RMSE and  $R^2$ .
- 69. How does train-test split works? How can they be used to validate the model.
- 70. What is cross validation techniques ? Explain the k-fold cross validation techniques with example.
- 71. Explain the need of cross validation techniques.
- 72. What are hyper parameter? Explain the need of hyperparameter tuning with example.
- 73. Explain the grid search and random search with example.