# Unit 1

1. Define Machine Learning. Explain with examples why machine learning is important.

2. Discuss some applications of machine learning with examples.

3. Explain how some disciplines have influenced the machine learning.

4. What is well- posed learning problems.

5. Describe the following problems with respect to Tasks, Performance and Experience:

   a. A Checkers learning problem

   b. A Handwritten recognition learning problem

   c. A Robot driving learning problem

   d. In Spam E-Mail detection

6. Explain the steps in designing a learning systems in detail.

7. Explain different perspective and issues in machine learning.

8. Define concept learning and discuss with example.

9. Explain the General-to-Specific Ordering of Hypotheses

10. Write FIND-S algorithm and explain with example given below and also find instance space, hypothesis space, and segmented distinct space.

Example 1:

| Example | Sky | AirTemp | Humidity | wind | water | forecast Water EnjoySport | |
|---|---|---|---|---|---|---|---|
| 1 | Sunny | Warm | Normal | Strong | Warm | Same | Yes |
| 2 | Sunny | Warm | High | Strong | Warm | Same | Yes |
| 3 | Rainy | Cold | High | Strong | Warm | Change | No |
| 4 | Sunny | Warm | High | Strong | Cool | Change | Yes |

Example 2 :

| Time | Weather | Temperature | Company | Humidity | Wind | Shopping |
|---|---|---|---|---|---|---|
| Morning | Sunny | Moderate | Yes | Mild | High | Yes |
| Evening | Rainy | Cold | No | Mild | Low | No |
| Morning | Sunny | Moderate | No | Normal | High | Yes |
| Evening | Sunny | Moderate | Yes | High | High | Yes |
| Morning | Cloudy | Cold | No | Mild | Low | No |
| Evening | Rainy | Moderate | Yes | High | High | Yes |
| Evening | Sunny | Moderate | Yes | Normal | Low | Yes |

Example 3:

| Origin | Manufacturer | Color | Decade | Type | Example Type |
|---|---|---|---|---|---|
| Japan | Honda | Blue | 1980 | Economy | Positive |
| Japan | Toyota | Green | 1970 | Sports | Negative |
| Japan | Toyota | Blue | 1990 | Economy | Positive |
| USA | Chrysler | Red | 1980 | Economy | Negative |
| Japan | Honda | White | 1980 | Economy | Positive |

11. What are the key properties and complaints of FIND-S algorithm?

12. Define Consistent Hypothesis and Version Space.

13. Write LIST-THEN-ELIMINATE algorithm.

14. Write the candidate elimination algorithm and illustrate with example.

15. What is Ensemble Learning? List and Explain Ensemble Learning Methods with Example.

16. Give difference between supervised learning and unsupervised learning.

17. Give difference between parametric method and non-parametric method in ML.

18. Write the final version space for the below mentioned training examples using candidate elimination algorithm and also find instance space, hypothesis space, and segmented distinct space.

Example – 1:

| Origin | Manufacturer | Color | Decade | Type | Example Type |
|--------|--------------|-------|--------|------|--------------|
| Japan | Honda | Blue | 1980 | Economy | Positive |
| Japan | Toyota | Green | 1970 | Sports | Negative |
| Japan | Toyota | Blue | 1990 | Economy | Positive |
| USA | Chrysler | Red | 1980 | Economy | Negative |
| Japan | Honda | White | 1980 | Economy | Positive |
| Japan | Toyota | Green | 1980 | Economy | Positive |
| Japan | Honda | Red | 1990 | Economy | Negative |

Example – 2:

| Example | Citation | Size | In Library | Price | Edition | Buy |
|---------|----------|------|------------|-------|---------|-----|
| 1 | Some | Small | No | Affordable | Many | No |
| 2 | Many | Big | No | Expensive | One | Yes |
| 3 | Some | Big | Always | Expensive | Few | No |
| 4 | Many | Medium | No | Expensive | Many | Yes |
| 5 | Many | Small | No | Affordable | Many | Yes |

Example – 3:

| Size | Color | Shape | Class |
|------|-------|-------|-------|
| Big | Red | Circle | No |
| Small | Red | Triangle | No |
| Small | Red | Circle | Yes |
| Big | Blue | Circle | No |
| Small | Blue | Circle | Yes |

Example – 4:

| Example | Citation | Size | In Library | Price | Edition | Buy |
|---------|----------|------|-----------|-------|---------|-----|
| 1 | Some | Small | No | Affordable | One | No |
| 2 | Many | Big | No | Expensive | Many | Yes |
| 3 | Many | Medium | No | Expensive | Few | Yes |
| 4 | Many | Small | No | Affordable | Many | Yes |

Example – 5:

| Example | Size | Shape | Color | Surface | Thickness | Target Concept |
|---------|------|-------|-------|---------|-----------|----------------|
| 1 | Circular | Large | Light | Smooth | Thick | Malignant (+) |
| 2 | Circular | Large | Light | Irregular | Thick | Malignant (+) |
| 3 | Oval | Large | Dark | Smooth | Thin | Benign (-) |
| 4 | Oval | Large | Light | Irregular | Thick | Malignant (+) |

Example 6:

| Example | Eyes | Nose | Head | Fcolor | Hair | Smile |
|---------|------|------|------|--------|------|-------|
| 1 | Round | Triangle | Round | Purple | Yes | Yes |
| 2 | Square | Square | Square | Green | Yes | No |
| 3 | Square | Triangle | Round | Yellow | Yes | Yes |
| 4 | Square | Triangle | Round | Green | No | No |
| 5 | Square | square | Round | Yellow | Yes | Yes |

19. Explain in detail the Inductive Bias of Candidate Elimination algorithm.
20. Explain Adaboost algorithm. Write Steps of Adaboost algorithm.
21. Compute final prediction using Adaboost ensemble methods.

| CGPA | Interactiveness | Practical Knowledge | Communication Skill | Job Profile |
|------|-----------------|---------------------|---------------------|-------------|
| >=9 | Yes | Good | Good | Yes |
| <9 | No | Good | Moderate | Yes |
| >=9 | No | Average | Moderate | No |
| <9 | No | Average | Good | No |
| >=9 | Yes | Good | Moderate | Yes |
| >=9 | Yes | Good | Moderate | Yes |

22. Compare the different types of machine learning.
23. What is Machine Learning? List out its applications and possible ethical issues of machine learning applications?
24. List out different issues in machine Learning.
25. Define: a. Supervised Learning b. Classification c. Regression

26. What is decision tree and decision tree learning?

27. Explain the difference between linear and logistic regression with examples.

28. List the advantages of SVM and how optimal Hyperplane differs from Hyper plane?

29. Explain K-nearest Neighbor algorithm.

30. Relate entropy and information gain.

31. Consider the training dataset given in the following table. Use weighted k-NN and determine the class. Test instance (7.6, 60, 8) and K=3.

| S.No. | CGPA | Assessment | Project Submitted | Result |
|-------|------|------------|-------------------|--------|
| 1 | 9.2 | 85 | 8 | Pass |
| 2 | 8 | 80 | 7 | Pass |
| 3 | 8.5 | 81 | 8 | Pass |
| 4 | 6 | 45 | 5 | Fail |
| 5 | 6.5 | 50 | 4 | Fail |
| 6 | 8.2 | 72 | 7 | Pass |
| 7 | 5.8 | 38 | 5 | Fail |
| 8 | 8.9 | 91 | 9 | Pass |

32. What is the role of kernels? Classify the different types of Kernel.

33. Distinguish between classification and Clustering.

34. Explain representation of the decision tree with examples.

35. Explain the concepts of Entropy and Information gain.

36. Describe the ID3 algorithm for decision tree learning with an example.

37. Give Decision trees for the following set of training examples

| Day | Outlook | Temperature | Humidity | Wind | PlayTennis |
|-----|---------|-------------|----------|------|------------|
| D1 | Sunny | Hot | High | Weak | No |
| D2 | Sunny | Hot | High | Strong | No |
| D3 | Overcast | Hot | High | Weak | Yes |
| D4 | Rain | Mild | High | Weak | Yes |
| D5 | Rain | Cool | Normal | Weak | Yes |

| D6 | Rain | Cool | Normal | Strong | No |
|-----|----------|------|--------|--------|-----|
| D7 | Overcast | Cool | Normal | Strong | Yes |
| D8 | Sunny | Mild | High | Weak | No |
| D9 | Sunny | Cool | Normal | Weak | Yes |
| D10 | Rain | Mild | Normal | Weak | Yes |
| D11 | Sunny | Mild | Normal | Strong | Yes |
| D12 | Overcast | Mild | High | Strong | Yes |
| D13 | Overcast | Hot | Normal | Weak | Yes |
| D14 | Rain | Mild | High | Strong | No |

38. Consider the following set of training examples :

(a) What is the entropy of this collection of training examples with respect to the target function classification?

(b) What is the information gain of a2 relative to these training examples?

| Instance | Classification | a1 |
|----------|----------------|-----|
| 1 | + | T |
| 2 | + | T |
| 3 | - | T |
| 4 | + | F |
| 5 | - | F |
| 6 | - | F |

39. Identify the entropy, information gain and draw the decision trees for the following set of training examples :

| Gender | Car ownership | Travel cost | Income Level | Transportation (Class) |
|---|---|---|---|---|
| Male | 0 | Cheap | Low | Bus |
| Male | 1 | Cheap | Medium | Bus |
| Female | 1 | Cheap | Medium | Train |
| Female | 0 | Cheap | Low | Bus |
| Male | 1 | Cheap | Medium | Bus |
| Male | 0 | Standard | Medium | Train |
| Female | 1 | Standard | Medium | Train |
| Female | 1 | Expensive | High | Car |
| Male | 2 | Expensive | Medium | Car |
| Female | 2 | Expensive | High | Car |

40. Discuss Hypothesis Space Search in Decision tree Learning.

41. Binary classification vs. Multiclass classification in machine learning.

42. What is Clustering? Explain K – Mean clustering in detail.

43. What is Dimensionality Reduction? Explain Dimensionality reduction in detail.

44. Difference between PCA and LDA.

45. Difference Feature Selection and Feature Extraction Methods.

46. Explain the difference between linear and logistics regression with example.

47. Discuss the decision tree algorithm in detail.

48. What are overfitting and underfitting in context of machine learning models?

   What is the major cause of overfitting and underfitting?

49. Consider the following confusion matrix of the win/loss prediction of cricket match. Calculate model accuracy and error rate, sensitivity, precision, F-measure and kappa value for the same.

|  | Actual Win | Actual Loss |
|---|---|---|
| Predicted Win | 85 | 4 |
| Predicted Loss | 2 | 9 |

50. While predicting malignancy of tumor of a set of patients using a classification model, following are the data recorded:

    (a) Correct predictions – 15 malignant, 75 benign

    (b) Incorrect predictions – 3 malignant, 7 benign Calculate the model accuracy, error rate, Kappa value, sensitivity, precision, and F-measure of the model.

51. Explain the process of K-fold-cross-validation method.

52. Explain K-mean clustering algorithm.

53. What is feature selection? Why it is needed? What are the different approaches of feature selection, briefly explain any one.

54. Explain Under-fitting and Over-fitting with an example.

55. What is Multiple Linear regression?

56. Elaborate the cross validation in training a model

57. What is model accuracy in reference to classification? Also Explain the performance parameters Precision, Recall and F-measure with its formula and example.

58. Show various distance-based similarity measure with its example.

59. Briefly explain K-Medoids

60. Differentiate PCA and LDA.

61. Write a note on KNN