Subject : Machine Learning

Class : III -I SEC A&B

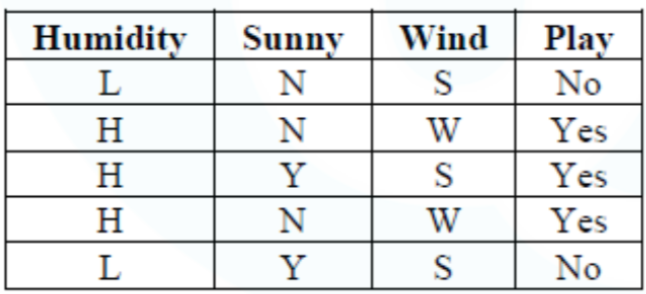
Department : Artificial Intelligence & Data Science

Note : All the questions are designed for 6M each.

**Unit – I**

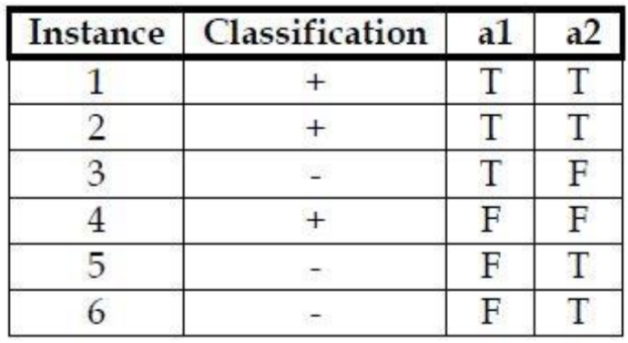
1. What is Machine Learning and list out its applications?
2. Define a well-posed learning problem. Discuss any 3 examples for well-posed learning problems along with its features.
3. Explain the steps in designing a Learning System in detail.
4. Differentiate between supervised, unsupervised and reinforcement learning.
5. Explain simple linear regression..
6. What is regression and explain the Multivariate Linear Regression
7. Formulate the types of regularization and explain how regularization is used to come out of overfitting
8. Differentiate between L1 and L2 regularization
9. Derive weight update equations for simple linear regression using gradient descent algorithm.
10. Write the Gradient Descent algorithm and differentiate between different variants of Gradient Descent algorithm
11. Explain the steps in designing a Learning System in detail.
12. Summarize the evaluation measures for regression techniques.
13. Discuss about k-fold cross validation

**Unit – 2**

1. Write about Logistic function along with its most convenient mathematical properties.
2. Explain the one-vs-rest approach of multi-class classification in Logistic Regression
3. Explain K-Nearest Neighbor learning algorithm
4. Write a short note on the following
   1. Distance measures used in K-Nearest Neighbors algorithm
   2. Normalization of data.
5. Discuss the Naive Bayes Classifier.
6. Consider the training data in the following table where Play is a class attribute. In the table, the Humidity attribute has values “L” (for low) or “H” (for high), Sunny has values “Y” (for yes) or “N” (for no), Wind has values “S” (for strong) or “W” (for weak), and Play has values “Yes” or “No”.
   1. 

What is class label for the following day **(Humidity=L, Sunny=N, Wind=W)**, according to naïve Bayesian classification?

1. Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute accuracy, precision, recall and F1 measure for the data
2. Explain the basic decision tree learning algorithm.
3. What do you mean by gain and entropy? How it is used to build the decision tree?
4. For the following set of training samples, find which attribute can be chosen as the root for decision tree classification.



1. Summarize the evaluation measures used for classification.
2. Identify the first splitting attribute for Decision Tree using Hunt’s algorithm with the following dataset.



**Unit – 3**

1. Define unsupervised learning. Discuss the applications of unsupervised learning.
2. What is clustering? Explain K-Means clustering algorithm.
3. Illustrate the K-means clustering algorithm and construct the K-means partitioning algorithm using the below dataset. Consider five points {A, B, C, D, E} with the following coordinates as a two-dimensional sample for clustering. A=(1,1); B-(1,0); C=(0,2); D=(2,4); E=(3,5)
4. Differentiate between agglomerative and divisive clustering. How the hierarchical clustering is represented?
5. Cluster the following 6 points using Single Link Hierarchical clustering.

|  |  |  |
| --- | --- | --- |
| Point | x | y |
| P1 | 0.40 | 0.53 |
| P2 | 0.22 | 0.38 |
| P3 | 0.35 | 0.32 |
| P4 | 0.26 | 0.19 |
| P5 | 0.08 | 0.41 |
| P6 | 0.45 | 0.30 |