Assignment 2 (Submission deadline: 31/8/2023)

1. A tourism company offers special discount card to its customers. Last year, they called many customers and a fraction of the customers accepted the offer. Here is the data that was collected by the Manager:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Serial No. | Job Type | Income Level | Likes to Hangout | Tours per year | Offer Taken |
| 1 | Engineer | High | Yes | 2 | Yes |
| 2 | Doctor | High | Yes | 1 | No |
| 3 | Engineer | Medium | No | 3 | Yes |
| 4 | Teacher | Medium | No | 2 | Yes |
| 5 | Doctor | High | Yes | 3 | Yes |
| 6 | Engineer | Medium | No | 2 | Yes |
| 7 | Teacher | High | Yes | 1 | No |
| 8 | Doctor | High | No | 1 | No |
| 9 | Teacher | High | No | 2 | Yes |
| 10 | Teacher | Medium | Yes | 3 | Yes |
| 11 | Engineer | High | No | 1 | No |
| 12 | Engineer | High | No | 2 | No |

Your task is to learn a Decision tree based on this data to predict whether a particular customer will take the offer or not. Among Income Level and Likes to Hangout which is the better attribute?

1. There are two boxes containing coins. The first box contains 60 gold coins and 40 silver coins. The second box contains 25 gold coins and 75 silver coins. One of the two boxes is randomly chosen (both boxes have probability 0.5 of being chosen) and then a coin is picked up at random from the chosen box. If a silver coin is picked up, what is the probability that it comes from the first box?
2. A survey has been done on students to assess their interest in hostel accommodation. The data obtained is as follows:   
   200 of them were male students. Among the male students, 80 were juniors (first and second year) with 50% interested in hostel accommodation and the rest were seniors with 70% interested in hostels. Among the 100 females, 70 were juniors with 60% interested in hostels and the rest were seniors with 80% interested in hostels.
   1. Based on this data, construct a full joint distribution among the three random variables Gender(G), Category(C) and Interest in hostel accommodation(H)
   2. Calculate the following probabilities from your table:
      1. Probability of a student being a junior
      2. Probability of a female student not being interested in hostels
3. The initial classifier for OR logic is randomly set to 0.5 X1 + 0.5 X2 - 1.25 = 0. Assuming the learning rate is 0.1 and the learner is using sigmoid activation function. What will be the updated equation after all the training examples are processed once?
4. The initial classifier for AND logic is randomly set to 0.75 X1 + 0.75 X2 - 1.25 = 0. Assuming the learning rate is 0.1 and the learner is using hard threshold function. What will be the updated equation after all the training examples are processed once?
5. What will be the prediction of PlayTennis in the following Table when Outlook = Sunny, Temperature = 20, Humidity = Normal and Wind = Strong using Naïve Bayesian classifier? For continuous attribute assume Gaussian distribution.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Outlook | Temperature | Humidity | Wind | PlayTennis |
| D1  D2  D3  D4  D5  D6  D7  D8  D9  D10  D11  D12  D13  D14 | Sunny  Sunny  Overcast  Rain  Rain  Rain  Overcast  Sunny  Sunny  Rain  Sunny  Overcast  Overcast  Rain | 40  45  42  28  18  15  12  26  16  27  25  29  38  24 | High  High  High  High  Normal  Normal  Normal  High  Normal  Normal  Normal  High  Normal  High | Strong  Strong  Weak  Weak  Strong  Strong  Strong  Strong  Weak  Weak  Strong  Strong  Weak  Strong | No  No  Yes  Yes  No  Yes  Yes  No  Yes  Yes  Yes  Yes  Yes  No |