Yeditepe University
Computer Engineering Department
CSE585 – Machine Learning
Spring 2022, Assignment-1
Due Date: 19-April-2022

1 Bayesian Networks*(25 points)*

Ali,Veli and Can are close friends who call each other regularly. Ali is interested in cooking and every evening he cooks. If Veli and Can call Ali at the same evening, the probability that Ali burns his food is 1/3. If one of them calls, the probability that Ali burns his food is 1/4. The probability that Veli calls Ali is 1/2 and the probability that Can calls Ali is 1/3. If there is no call, Ali can cook without burning his food.

a) What is the probability of an evening that both Veli and Can calls Ali, but he does not burn his food?

b) What is the general probability for Ali to burn his food in an evening?

Draw the Bayesian belief network that describes the dependencies between the events. Show your probability assignments and steps.

2 Decision Tree Learning*(25 points)*

a	b	c	d	e	f	g	h	y
0	0	0	0	0	0	0	0	**0**
0	0	0	0	0	0	0	1	**1**
0	0	0	0	0	0	1	0	**0**
0	0	0	0	0	0	1	1	**1**
...	...	...	...	...	...	...	...	...
1	1	1	1	1	1	1	1	**1**

Table 1: A binary dataset

Consider the binary dataset given in Table 1. The dataset consists of 8-attributes $(a-h)$ and y column denotes the class label of each instance. Assume that the dataset consists of all possible 256 instances and the class label is the same with the last attribute h. In this case what will be the decision tree that would be generated by the decision tree learning algorithm?

Now assume that an error is introduced to the dataset and randomly chosen 25% of the instances have the class label y set to the opposite of h. Now assume that these 256 instances are divided into a training and a test set randomly. What can you say about the decision tree that would be generated? *(You do not need to draw the tree!)* What would be the success rate of this decision tree on the training set?

More importantly, can you calculate the accuracy of the tree on the test set?

So what is the name of the problem that is introduced with this corrupted data?

What is the solution to the problem?

3 Concept Learning *(25 points)*

Consider a company which exports and imports goods. Consider the following positive and negative training examples for the items that can be ordered at the same time. Each example shows two items that can be ordered at the same time where each order is described with the attributes *Transfer Type (export, import)*, *Destination (Europe, Asia, Africa)*, *Quantity (high, medium, low)* and *Type (automotive, food, textile, metal, chemicals, cement, equipment)*.

```
+  <<export,Asia,high,automotive><import,Europe,low,automotive>>
+  <<export,Asia,low,food><import,Europe,low,automotive>>
-  <<import,Asia,high,textile><import,Europe,low,chemicals>>
+  <<export,Asia,high,metal><import,Europe,low,metal>>
```

a) Provide the version space that would be formed by the *Candidate Elimination* algorithm on this dataset. Show your steps!

b) If we had only the following positive training example, how many hypothesis in the hypothesis space would be consistent with this single training example?

```
+  <<export,Europe,low,equipment><import,Africa,high,chemicals>>
```

4 Genetic Programming*(25 points)*

Consider the 4-Op problem where you are given a set of integers and a target integer. The aim is to find an expression that consists of the four arithmetic operators and the given set of integers where the result of the expression would be equal to the target integer.

For instance if the given numbers are $3, 6, 7, 8$ and the target integer is 19, a possible solution is $((3 * 6) + (8 - 1))$. Now assume, you apply genetic programming to solve the problem. Would you encounter a problem when you apply the standard crossover and mutation operations on the chromosomes in the population? Define this potential problem and try to design a solution to overcome the difficulty.