**EECS 349 Homework 4**

**Wenting Zhou NetID:wzg249**

***Question 1:***

The dataset I created contain 30 attributes (X1~X3 are numeric and the others are nominal) and one label Y (nominal: 0 or 1). The creation is based on four rules:

1. If x1 < 5 and x7 = 0, then output Y is 1;
2. If x10=1 and x9= 0 and x2=0, then output Y is 1;
3. If x8=1 and x6=1, then output Y is 1;
4. If x12=0 and x21=0 and x17=1, then output Y is 1.

For the data do not meet these rules, the output Y is randomly picked as 0 or 1. Because these rules can be easily got by decision tree and hard to learn by nearest neighbor, the testing results based on 10-fold cross-validation show 99.3% accuracy for J48 and 65% accuracy for IBk.

***Question 2:***

The dataset has two attributes x1, x2 (nominal) and one output Y (nominal). The creation follows the rule (XOR) that if x1 and x2 has the same value then the output should be 0, and if not the output should be 1. This rule can be perfectly learned by Multilayer Perceptron with 100% accuracy. However, BayesNet performs bad with only 52.9% accuracy, because there is no probability dependence among the attributes and output.

***Question 3:***

1. Genetic Algorithms

*Example:* minimize the value of function f(x) where f(x)=2a+6b+17c+3d+9e-15 and a, b, c, d, e have different integrate values ranging from 0 to 30.

Here we use genetic algorithm and get 6 chromosomes in the form [a, b, c, d, e]:

Chromosome [1] = [22, 01, 23, 24, 16]

Chromosome [2] = [21, 10, 02, 25, 11]

Chromosome [3] = [20, 30, 03, 29, 12]

Chromosome [4] = [04, 09, 26, 15, 17]

Chromosome [5] = [28, 05, 08, 13, 18]

Chromosome [6] = [07, 27, 06, 14, 19]

Then we can do the calculation and crossover again and again until we find the best result. If we reduce the number of chromosome, then we should include mutation in order to try each value from 0 to 30.

*Reasons:* The function and problem of this example meet the features of genetic algorithm and reduce the calculation. For instance, the values for parameters are integrate and discrete which can be list in the form like chromosome. The crossover and mutation can realize all the combination of a 5-number group. Considering the number of unknown parameters, hill climbing may block in local minima and need a large amount calculation. In terms of gradient descent, it is inconvenient to calculate gradient for 5 parameters and the discrete data do not offer gradient descent approach any advantage in this example.

1. Gradient descent

*Example:* find the values for and to minimize the value of function f(x) where (all the parameters are continuous numeric number)

*Reasons:* Here we use gradient descent for the following reasons. First, the parameters are continuous which is suitable for gradient descent but not for hill climbing. Second, it is not difficult to do the differentiation for and , and the formulas of differentiation are easy to calculate. Third, this example problem has convexity and gradient descent is the ideal approach in this situation. We do not use hill climbing because it is more suitable for discrete data. And the reasons for not using genetic algorithm is that continuous data can not form the chromosome units, and a large number of chromosome, crossover and mutation is required in this example. In another word, the number of parameters are too small to form a meaningful model of chromosome.

1. Hill climbing

*Example:* find one local maximum branch in a tree structure, which means the total value of passing nodes between root node and final leaf reaches its maximum.

Reasons: First, the structure of data is not convexity and the values of data are not continuous which make gradient descent unsuitable. If the tree is not a full tree, the missing nodes may make the path from root nodes to final leaf contain different numbers of passing nodes which may add difficult to create chromosomes units for genetic algorithm. Hill climbing are easily applied in this question and there is no need to find the global maximum which avoids the shortage of hill climbing and makes it the ideal approach for this example.

***Question 4***:



The minimum number of probabilities is 8+1+1+2+4+1=17

1. The number of independent parameters needed to specify the full joint distribution over the six variables is 17.
2. G, A, and E form a head-to-head structure, and A blocks G and E when A is not known, which means if A is unknown, G and E are inactive connected and independent.
3. In the head-to-head structure, G and E are dependent given A.

***Question 5:***