Report on Assignment 4

1] Independent Bayesian networks:

The test set log-likelihood of the Bayesian network with no edges for the 10 datasets is given below

Datasets	Log-Likelihood
accidents	-65.728
baudio	-71.178
benetflix	-93.142
dna	-144.826
jester	-92.163
kdd	-3.528
msnbc	-9.767
nltcs	-13.321
plants	-45.107
r25	-162.484

Table 1: Independent Bayesian networks (Log-Likelihood)

2] Tree Bayesian networks:

The test set log-likelihood of the tree Bayesian networks by using Chow-Liu algorithm for the 10 datasets is given below:

Datasets	Log-Likelihood
accidents	-47.834
baudio	-64.021
benetflix	-86.884
dna	-126.039
jester	-83.944
kdd	-3.313
msnbc	-9.544
nltcs	-9.747
plants	-23.816
r25	-139.857

Table 2: Tree Bayesian networks (Log-Likelihood)

3] Mixtures of Tree Bayesian networks using EM:

The test-set log-likelihood of the mixtures if trees using EM algorithm for the given 10 datasets is given below.

Datasets	K-value	Mean	Standard Deviation
accidents	20	-44.322	0.103
baudio	15	-58.455	0.027
benetflix	20	-82.015	0.101
dna	15	-119.325	0.121
jester	15	-75.869	0.012
kdd	5	-3.305	0.014
msnbc	15	-9.112	0.006
nltcs	20	-8.634	0.003
plants	15	-19.095	0.056
r25	20	-125.835	0.199

Table 3: Mixtures of Tree Bayesian networks using EM (Log-Likelihood)

Here the mean and standard deviation is calculated by running the algorithm for 10 times on each dataset and the values of K is decided by using the validation set of the respective datasets.

4] Mixture of Tree Bayesian networks using random forest:

The test-set log-likelihood of the mixture of tree Bayesian network using random forest for the given 10 datasets is given below.

Datasets	K - values	r - values	Mean	Standard Deviation
accidents	20	40	-47.831	0.002
baudio	15	35	-64.062	0.014
benetflix	15	45	-86.899	0.0005
dna	10	50	-126.726	0.009
jester	15	35	-84.269	7.02e-05
kdd	10	30	-3.314	0.001
msnbc	10	5	-9.399	0.0001
nltcs	20	5	-9.712	0.002
plants	20	30	-23.798	0.001
r25	15	50	-139.886	0.004

Table 4: Mixture of Tree Bayesian networks using random forest (Log-Likelihood)

Here the values of K and r is selected by experimenting the data with its validation set, also the mean and standard deviations is calculated by running the algorithm for 10 times in their respective test sets.

From above experiment it seems that mixture of tree Bayesian networks using EM algorithm performed well, because it uses concept of chow-liu with soft clustering. The chow-liu method and the random forest technique provides nearly same results. But the independence Bayesian networks performed worst from all other parts in this experiment.