MACHINE LEARNING HW4

Part 1

Tree Bayesian networks.

Dataset File	Log Likelihood
Accidents	-33.02869130872363
Baudio	-44.25205757092208
Bnetflix	-60.146803791038984
Jester	-58.194533381241776
Msnbc	-6.5398377728276325
Nltcs	-6.760056867087026
Plants	-16.222419605938835
Pumsb_star	-30.82360635324834
Tretail	-10.810237548877748
kdd	-2.5284779278146976

Part 2

Mixtures of Tree Bayesian networks using EM

Best Parameters with their Average Log Likelihood and Standard Deviation-

Dataset File	k	Max_iter	AVG Log Likelihood	Standard Deviation
Accidents	20	50	-29.5979088424769	0.030483833912594
Baudio	20	50	-40.0262869198352	0.02550171638944204
Bnetflix	20	50	-56.7153276907902	0.044876665530108564
Jester	10	50	-53.092943874864	0.045739493436862
Msnbc	20	50	-6.146630523238571	0.002456401605482711
Nltcs	20	50	-6.016386129156	0.00170154865868
Plants	20	50	-13.13413638194	0.030329264963422
Pumsb_star	20	50	-23.67928242254008	0.03780284820349636
Tretail	5	50	-10.82352047174108	0.004028776890526212
kdd	20	50	-2.1449734165917	0.0035609534765529

Part 3

Mixtures of Tree Bayesian networks using Random Forests.

Best Parameters with their Average Log Likelihood and Standard Deviation-

Dataset File	k	r	AVG Log Likelihood	Standard Deviation
Accidents	200	10	-32.958053482888936	0.0019271263830096628
Baudio	200	20	-43.69537669426694	0.002260060825241342
Bnetflix	200	50	-59.82153251305824	0.006300230259570801
Jester	100	50	-57.3215880893849	0.01052893797144711
Msnbc	10	5	-6.53769007601606	0.00028727121133747
Nltcs	50	15	-6.69986115394631	0.0057740991574656
Plants	50	40	-16.0072785877125	0.01233833660528232
Pumsb_star	50	10	-30.7173552181631	0.00712758028170989
Tretail	200	10	-10.7508080770975	0.0005046816853063
kdd	100	20	-2.494422189219	0.0006796698358634

- Can you rank the algorithms in terms of accuracy (measured using test set LL) based on your experiments? Comment on why you think the ranking makes sense.

  The algorithms can be ranked in the following order-

  1. Mixtures of Tree Bayesian networks using EM
  2. Mixtures of Tree Bayesian networks using Random Forests
  3. Tree Bayesian networks

  EM algorithm makes use of Mixture probabilities that give good performance over Random Forests. The Average Log likelihood of the Tree Bayesian Network (Chow Liu Tree) is poor over the other two as only one tree is learnt.