Homework 4
Machine Learning

Due date: Sunday, April 29, 11:59 p.m.

In this project, you will implement the following four algorithms and test their performance on 10 datasets available on the class web page.

1. **Independent Bayesian networks.** Assume the following structure: The Bayesian networks has no edges. Learn the parameters of the independent Bayesian network using the maximum likelihood approach. Use 1-Laplace smoothing to ensure that you don't have any zero probabilities in the model.

2. **Tree Bayesian networks.** Use the Chow-Liu algorithm we discussed in class to learn the structure and parameters of the Bayesian network. Use 1-Laplace smoothing to ensure that you don't have any zeros when computing the mutual information as well as zero probabilities in the model.

3. **Mixtures of Tree Bayesian networks using EM.** The model is defined as follows. We have one latent variable having k values and each mixture component is a Tree Bayesian network. Thus, the distribution over the observed variables, denoted by $\mathbf{X}$ (variables in the data) is given by:

$$P(\mathbf{X} = \mathbf{x}) = \sum_{i=1}^{k} p_i T_i(\mathbf{X} = \mathbf{x})$$

where p_i is the probability of the i-th mixture component and T_i is the distribution represented by the i-th Tree Bayesian network. Learn the structure and parameters of the model using the EM-algorithm (in the M-step each mixture component is learned using the Chow-Liu algorithm). Select k using the validation set and use 1-Laplace smoothing. Run the EM algorithm until convergence or until 100 iterations whichever is earlier.

4. **Mixtures of Tree Bayesian networks using Bagging.** The model is defined as above (see (3)). Learn the structure and parameters of the model using the following Bagging-style approach. Generate k sets of Bootstrap samples and learn the i-th Tree Bayesian network using the i-th set of the Bootstrap samples (as before use the Chow-Liu algorithm to learn the structure and parameters of the Tree Bayesian network). Select k using the validation set and use 1-Laplace smoothing. You can either set $p_i = 1/k$ for all i or use any reasonable method (reasonable method is extra credit). Describe your (reasonable) method precisely in your report. Does it improve over the baseline approach that uses $p_i = 1/k$.

Report Test-set Log-Likelihood (LL) score on the 10 datasets available on the class web page. For EM and Bagging (since they are randomized algorithms), run the algorithm 10 times on each dataset and report average test set log-likelihood and the standard deviation over the 10 runs. 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.

What to turn in:

- Your report in PDF format describing your experimental evaluation.
- Code along with a Readme file on how to compile and run your code.
- **Note:** Please submit a single zip file containing your report and your code (rar/gzip/tar and other formats are not allowed).