Problem Set 5

CS 6347

Due: 4/23/2017 by 11:59pm

Note: all answers should be accompanied by explanations for full credit. Late homeworks cannot be accepted. All submitted code **MUST** compile/run.

Problem 1: Bayesian Networks of Student Performance (100pts)

For this problem, you will use the `student_train.data` and `student_test.data` data sets provided with this problem set. This data set was generated from the UCI Student Performance data set by pruning the attributes from around 30 to around 20 (to simplify your lives). We will be interested in predicting student performance (the last column) given the other attributes. You should include your MATLAB code for both subproblems in your submission.

Naïve Bayes (40pts): Train a naïve Bayes classifier to predict the end of semester grade using `student_train.data`. What is your accuracy on `student_test.data`? Is this performance reasonable?

Structure Learning (60pts): Learn a discriminative Bayesian network for the 'grade' attribute by using the Chow-Liu Bayesian structure learning algorithm described in class. For this problem, you should turn in your directed tree and MATLAB code. Note: MATLAB has toolbox support for finding a minimum spanning tree. How does your learned structure compare with that of naïve Bayes?