Problem Set 5

CS 6347

Due: 4/21/2019 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 Learning (100pts)

For this problem, you will use the `mushroom_train.data` and `mushroom_test.data` data sets provided with this problem set. This data set was generated from the UCI Mushroom. We will be interested in predicting edibility of the mushroom (the first 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, i.e., a tree structure with the label at the root and all other random variables as leaves, to predict whether or not the mushroom is poisonous using `mushroom_train.data`. What is your accuracy on `mushroom_test.data`? Is the performance reasonable? Explain how to improve performance using a prior.

Structure Learning (60pts): Learn a discriminative Bayesian network for the class label (the first column) 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 both with and without regularization?