Instructions for Experimenter

**Introduction to Bayes Nets**

Bayes nets are a type of machine learning model that allow a user to make predictions about the probability of a particular event occurring based on previous events. In this tutorial, we will study the behavior of a student attending an 8 AM tutoring session while familiarizing ourselves with the *Bayes Net Debugger* program. The program allows you to build a Bayes Net graph to represent your machine learning model, then run examples through the graph to see how well your model classifies test data points.

**The Problem**

Alice is a freshman at MIT who needs extra tutoring for 8.01 (physics). However, the only time the tutor is available to meet is at 8 in the morning on Sundays. Attending sessions at 8 AM may be tough for Alice for several reasons: she often goes out partying Saturday nights and thus goes to bed really late, she is not a morning person and is often too lazy to get out of bed at an early time, she is taking a rigorous schedule at MIT so sometimes she would rather do other work than attend tutoring sessions. Our job is to help the tutor figure out the probability that Alice will attend the tutoring session because the tutor does not want to get out of bed at 8 AM if she doesn’t have to. We can figure this out by looking at how Alice behaved in past situations and use conditional probabilities based on other events to calculate the probability that Allice will attend the next session.

**Tutorial**

To solve this problem we will first build a graphical representation of the problem, and then run the test data through the graph. The goal is to achieve the highest test accuracy possible. **Follow the instructions in bold** in order to familiarize yourself with *Bayes Net Debugger*:

1. **Upload the data to the program.** Click “choose file” at the bottom left on the program and select the appropriate data file (AliceData.csv). Then click the “Upload Data” to upload the data to the program.
2. **Create a node and rename it “Attended 8AM Session”.** Drag the oval-shaped node from the toolbar into the graph-building section (big square). Double click the node to rename it as a feature.
3. **Create a second node and name it “Partied Night Before”.**
4. **Highlight a node by Ctrl+clicking the node** (to unhighlight press Ctrl+click again).
5. In order to view how the data looks with this graphical model, we can animate the nodes. The nodes will change color (from white) if the feature is true at that data point. The nodes will turn the same color if all features are true. This is helpful because it will allow you to visualize the data. If one feature only seems to be true when another feature is true, then it is likely that there is some correlation between the two features. To animate nodes, highlight the chosen nodes and then click “Start Animation”. You can even select the speed at which the colors flash by sliding the bar below the “Stop Animation” button. **Animate the nodes “Attended 8 AM Session” and “Partied Night Before”. Do you suspect that the two events are correlated to one another?**
6. Connect nodes on the graph by clicking the “Connect Mode” button. Connect mode is ON when the “connect mode” button is gray and OFF when the “connect mode” button is white. Once in connect mode, select the parent node by clicking on one node, then select the child node by clicking on another node. **Connect the nodes “Partied Night Before” and “Attended 8AM Session”, with “Partied Night Before” being the parent node. Exit connect mode when finished.** (To disconnect nodes**,** highlighting the two nodes you want to disconnect, then pressing the “Disconnect Highlighted Nodes” button).

Testing:

1. **At the top of the application next to the “TESTING” title, set the “Select Final Node” Dropdown to be “Attended 8AM Session”**. This is the feature we ultimately want a value for.
2. Testing with manually inputted values: this module allows you to specify values for nodes in the graph. It will then, for each node without an assigned value, calculate the probability that the node’s value is true and randomly assign a value to that node based on the probability. **Assign a value of true to the “Partied Night Before” feature and a value of “not given” to the “Attended 8AM Session” feature. Click “Assign Values” to step through the program and get a value for each node.** **Click “clear results: to clear all values from the nodes. Try using this module again except this time set “Partied Night Before” to false. Observe how the probability of being true changes for the “Attended 8AM Session”.**
3. Now we will test the accuracy of our model with testing data. Focus your attention to the “Test with Testing Data” section. **Choose the “AliceTest.csv” file and click “Upload Test Data”.** **You can quickly run through all tests at once (and select the speed at which you do so by sliding the bar below) by clicking “Run All Tests”. Or you can step through the test points one at a time using the “Single Step Through Tests” button**. What is the test accuracy? Do you think it can be improved?
4. If the test accuracy does not appear to your liking, you can also split and merge nodes. **Try splitting the node “Partied Night Before” into two nodes by highlighting the partying node and clicking “Split Highlighted Nodes”. Rename the split nodes to “Drinking” and “Not Drinking”. See if this improves your results.**
5. If splitting the nodes does not improve your results, merge the two nodes back together by highlighting them and clicking “Merge Two Highlighted Nodes”.
6. **Try adding more nodes to see if you can increase the test Accuracy.**

**The Second Problem**

Now that you have a feel for how the program works, we present you with a second problem:

There are many factors that determine whether a plane is able to land at its destination. Some factors are the weather, and the airline’s ability to stay organized. Imagine that you have a friend who is supposed to land at Boston Logan Airport. You need to pick your friend up from the airport, so you want to know ahead of time whether or not his/her plane will land. Unfortunately, the airline doesn’t give you the status of the plane, so you must pay attention to other factors in order to make your decision. To make your decision, you will build a Bayes Net. An initial Bayes net is built for you (shown below). Training data is provided that will allow you to compute the probability of the plane landing based on other events. Please look at the file, planeLandData.csv for training data and testData.csv for test data.