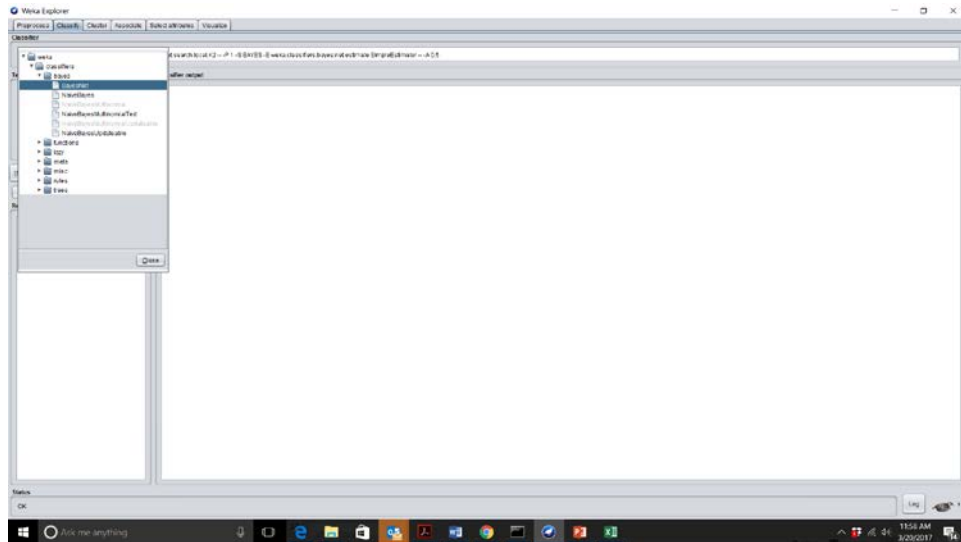


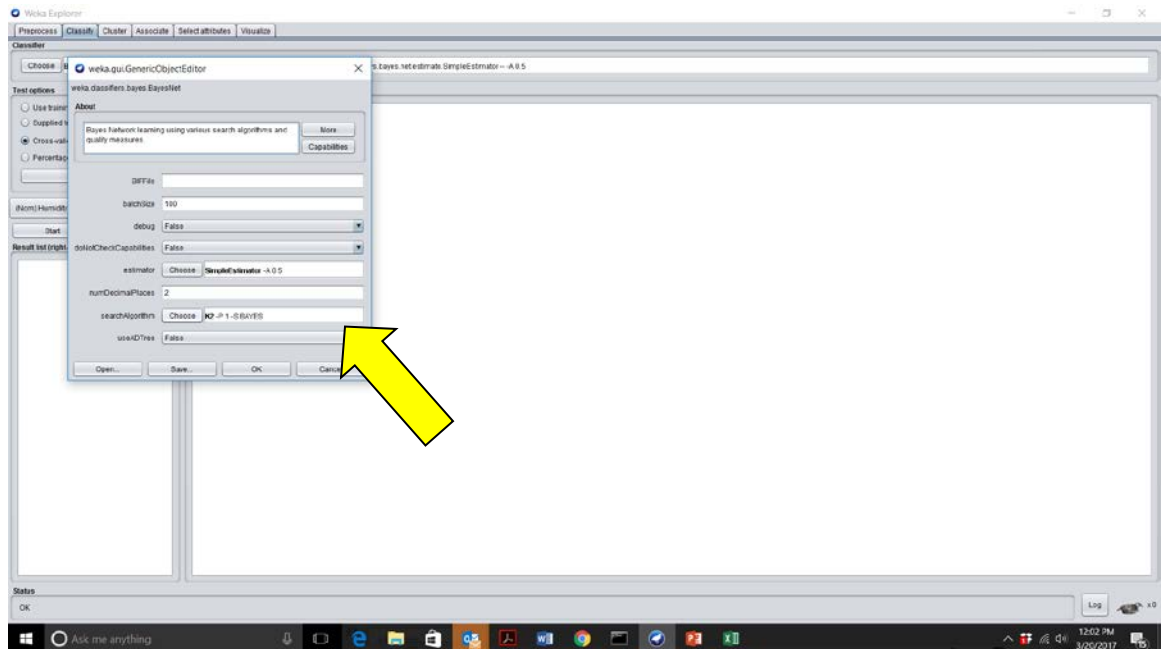
# Data Mining II: Advanced Methods and Techniques

## Assignment #3: Bayesian Networks

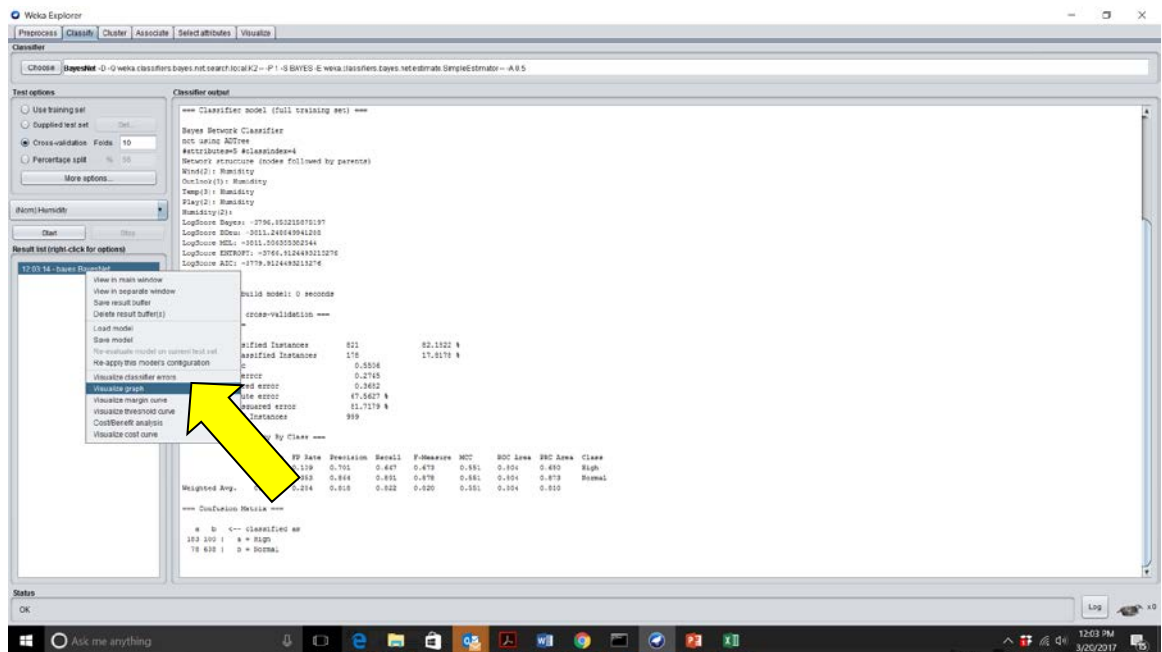
1. What is Bayes Rule Formula?
2. What does “maximum likelihood” means?
3. What is Bayesian learning?
4. What is a prior? What is posterior?
5. In Weka open the “BayesNetwork\_Weather\_data\_Set.csv” and under the “Classify” Tab click on the Bayes Folder and then BayesNet method.



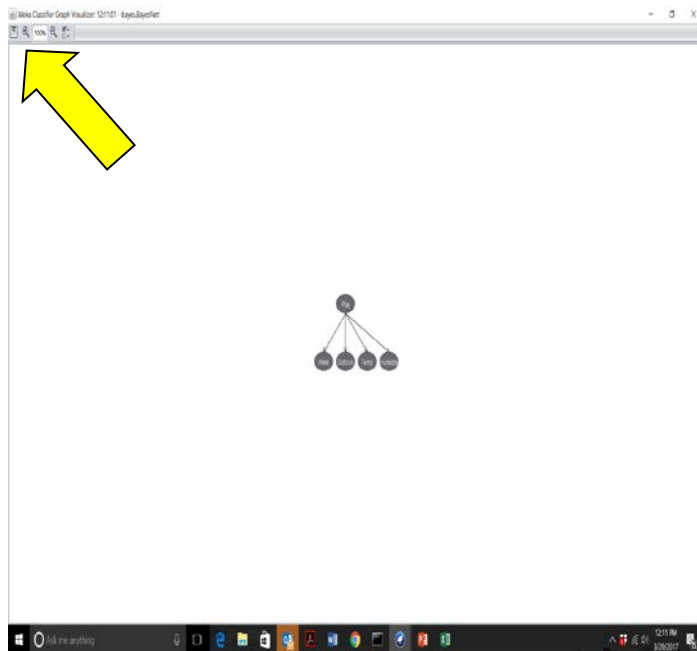
Train the Bayesian NEtowrk on the data set using several different parameters. First Train the netowrk by applying the Simple K2 search algorithm



You can visualize the model as well as the graph:



Once you click on visualize the learned graph structure :

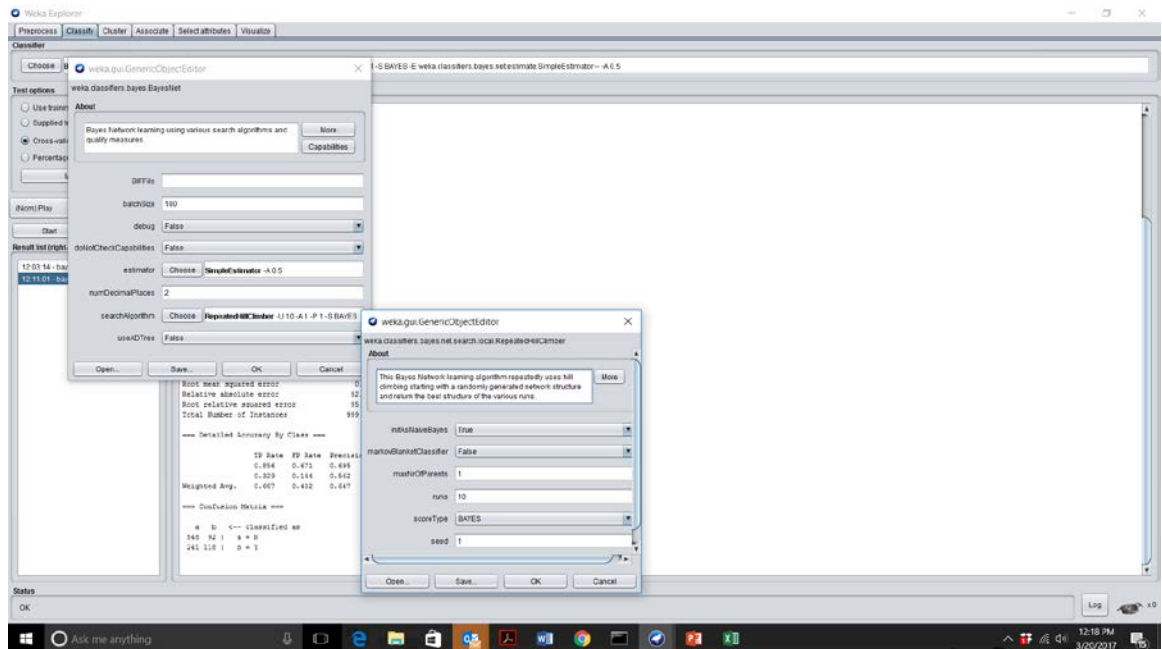


Save the graph in the XML BIF file format

You can double click or right click on the nodes and visualize the probabilities.

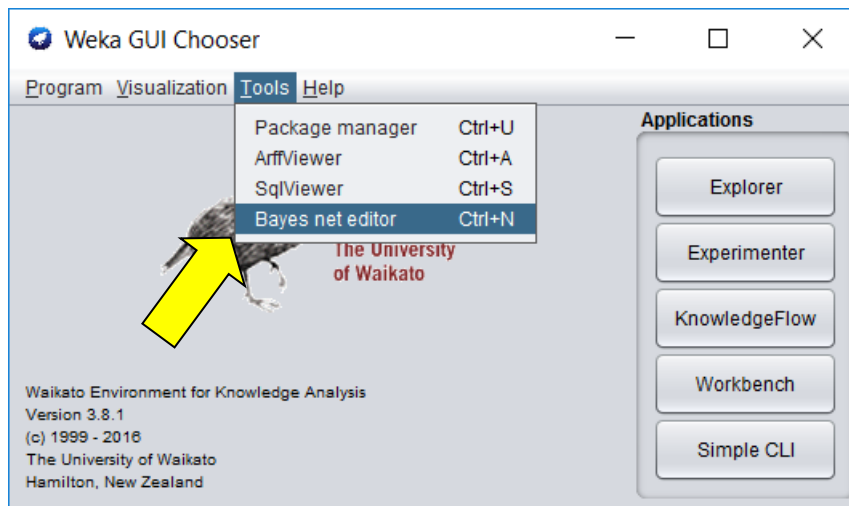
Repeat the training process by utilizing different search algorithms (ReapedHillClimber, SimulatedAnnealing, etc under BayesianNet parameters options). Save each graph, compare and contrast the confusion matrix, output classification errors and predictive power of each model. Describe and discuss in details the differences, pros/cons between the models.

Each of the searching algorithms will have a set of their own parameters you can adjust:



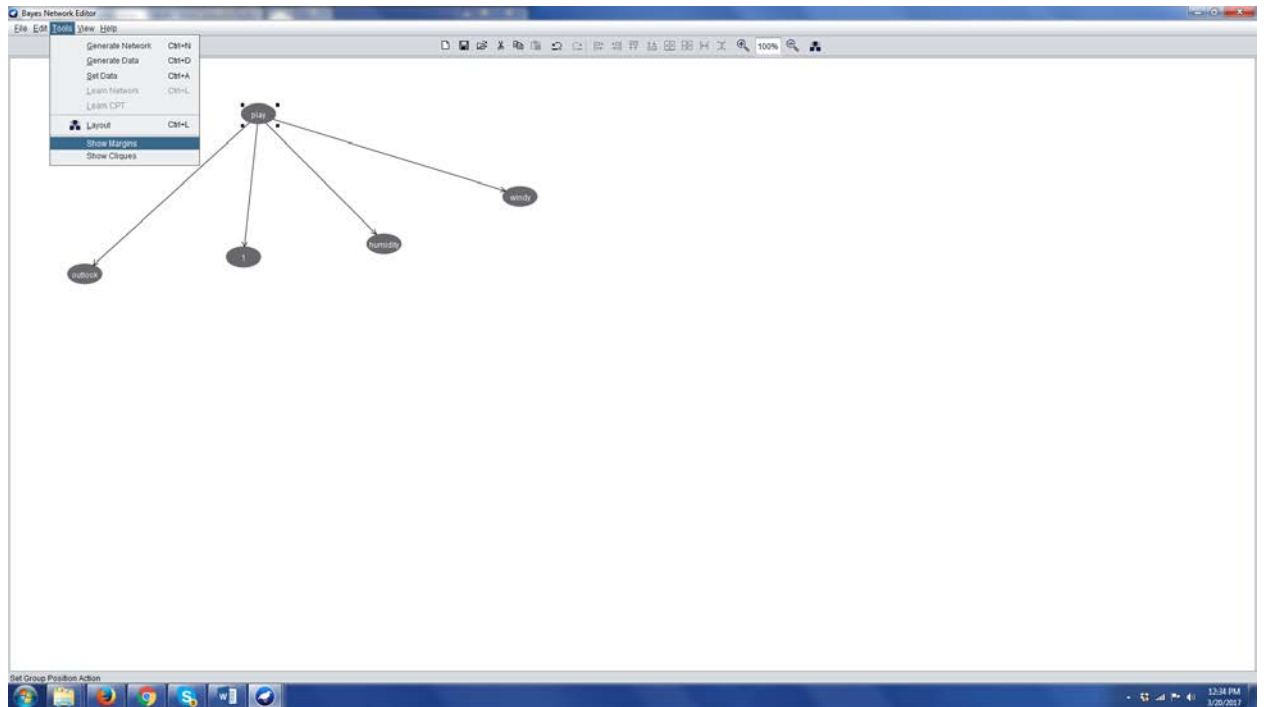
Part 2:

From Weka's Start Menu click on Tools and then Bayes net Editor

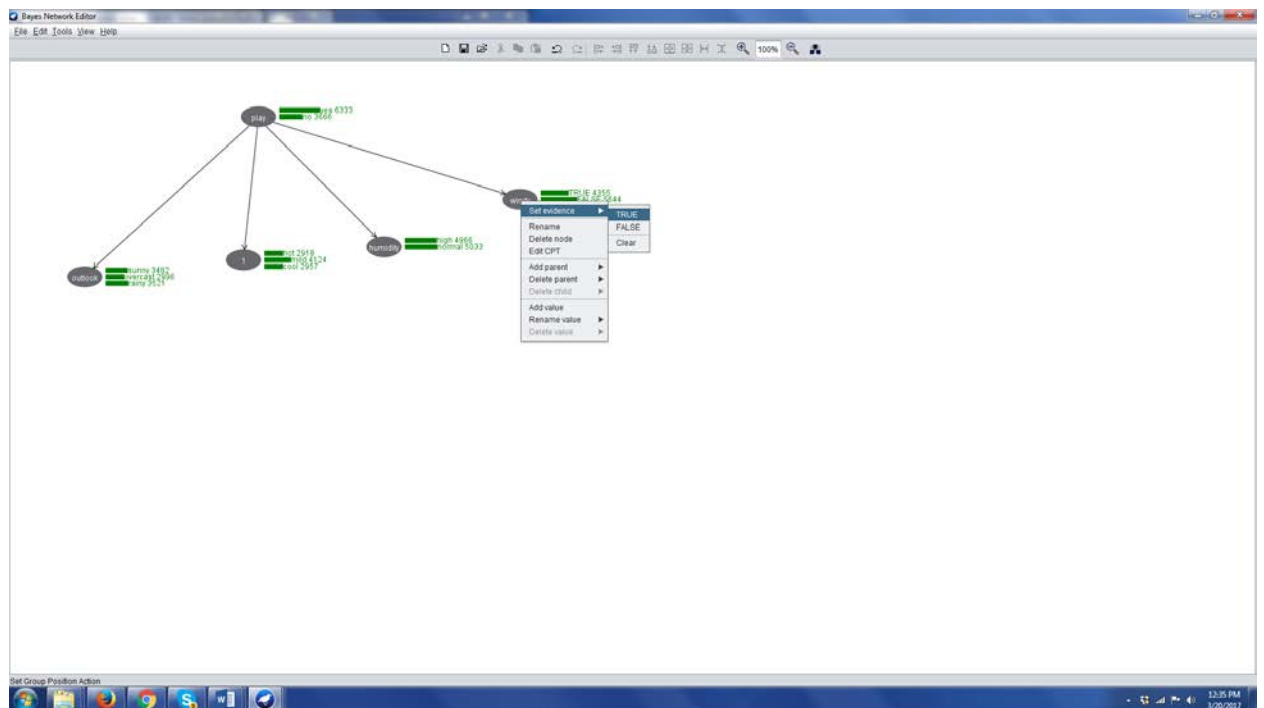


Now you can load in the graph structure you have previously saved in the XML BIF format.

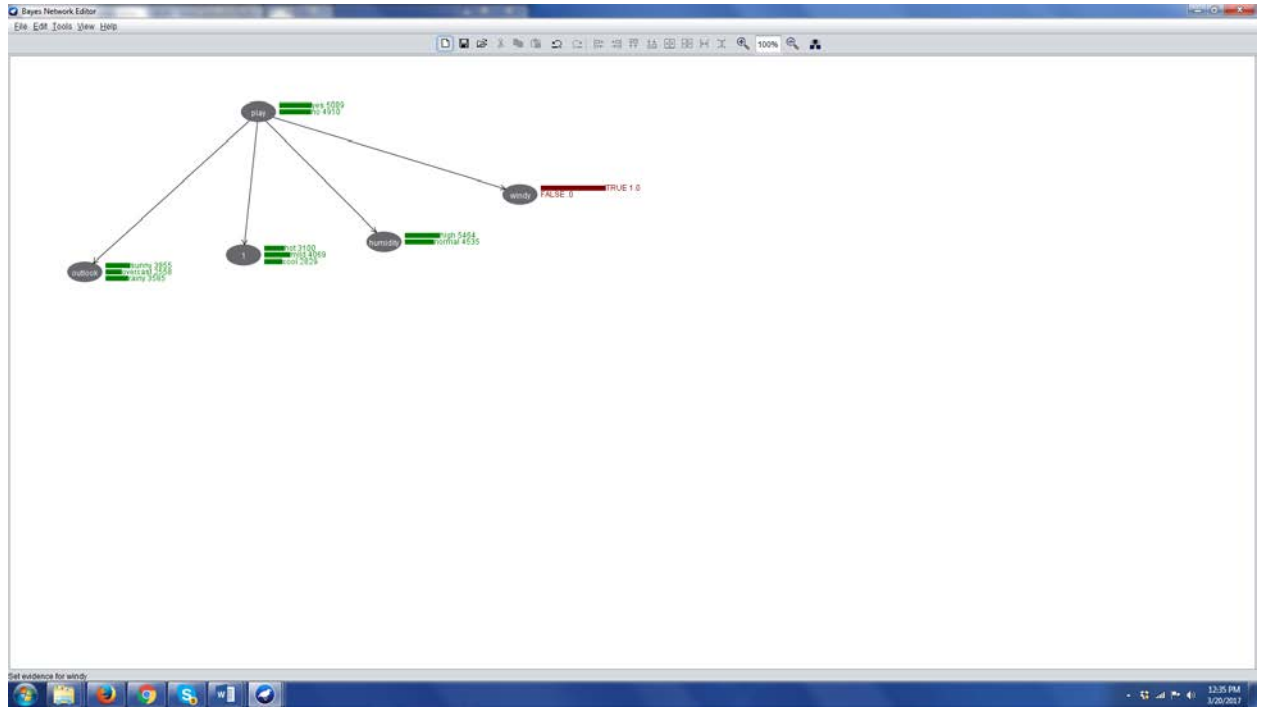
In this Editor, you can now click on Tools – Show Margins to see the probabilities associated with probability table for each node.



We can now set evidence on any of the nodes:



Notice how the rest of the node's probabilities change:



Describe and discuss the apparent changes. Observe how the evidence has propagated, what else changed? Why?

In this Editor, you can delete or add arcs and/or nodes from the Edit tab. Once you have deleted the arc of your choice, you can retrain by set Data and Learn CPT commands. You can now set evidence on one of the nodes and explain how these changes influenced the posterior probabilities in the graph.

6. Follow the process for training and visualizing the BN graph from the Question #5. Train the Bayesian Network and Decision Tree model on Hypothyroid data set. Describe the process and evaluation of the models. What are pros and cons of each? Which one performed better? Why? Which model is more suitable for particular applications?