CIS 730 Artificial Intelligence
CIS 530 Introduction to Artificial Intelligence
Fall 2018

Homework 8 of 10: Machine Problem
Reasoning and Learning, Part II:
Probabilistic Reasoning (Inference and Causality),
Version Spaces, and Decision Trees; More WEKA

Assigned: Sun 04 Nov 2018
Due: ~~Mon 05 Nov 2018~~ Sun 11 Nov 2018 (before **midnight**)

The purpose of this assignment is to exercise your knowledge of graphical models of probability and basics of machine learning. You will solve some Bayesian network reasoning and learning tasks and a few classification problems to simulate the behavior of supervised inductive learning algorithms, and continue working with classification models in WEKA.

This homework assignment is worth a total of 100%.
Each problem is worth 25% for CIS 530 students and 20% for 730 students.

References

***TETRAD*:** http://www.phil.cmu.edu/projects/tetrad/
Hugin Lite: http://www.hugin.com/index.php/hugin-lite/
Hugin Tutorials (PDF): http://bit.ly/hugin-tutorials-pdf
WEKA: http://www.cs.waikato.ac.nz/ml/weka/
***WEKA* Data:** http://www.cs.waikato.ac.nz/ml/weka/datasets.html
Version Spaces: http://www.emunix.emich.edu/~sverdlik/DM/VersionSpace.html

1. **(530/730) Hugin Tutorials, Concluded. (530/730) Hugin.** If you have not already done so (due to skipping MP7), download Hugin Lite for your platform from http://www.hugin.com/index.php/hugin-lite/, install it, and use it to construct the influence diagram (*aka* decision network) for Tutorial 2 (the Apple Tree), following the instructions in **http://bit.ly/hugin-tutorials-pdf.** Turn `ps8_1.zip`, an archive containing screenshots of your demo as you step through the construction and running of the model. **What kind of temporal Bayesian network representation does the network implement?**

2. **(530/730) TETRAD Structure Learning.** Use TETRAD to learn the structure of a Bayesian network using any categorical (discretized or nominal) data from the UCI Machine Learning Database Repository (see the WEKA data set page: http://www.cs.waikato.ac.nz/ml/weka/datasets.html). **Hint:** focus on using data for clustering tasks. Turn in `PS8-2.pdf` or include your solution within a single `PS8.pdf` file, consisting of screenshots of the network learned.

3. **(530/730) Bayesian Network Inference.** In your own words, differentiate between exact and approximate inference, and describe at least two exact inference algorithms for general multiply-connected networks (e.g., loop cutset conditioning; *Hugin* or junction tree, also known as the Lauritzen-Spiegelhalter algorithm; and variable elimination). Compare these to at least one approximate inference algorithm (e.g., forward simulation). Discuss whether and how this algorithm would work on the network structure you learned in PS8-1. Turn in your solution as `PS8-3.pdf` or as part of `PS8.pdf`.

4. **(730 only) Causal Discovery.** Watch this year's video selection posted in Canvas and discuss how Bayesian network structure learning and parameter estimation might or might not work for causal discovery using the data described. Turn in your solution as `PS8-4.pdf` or as part of `PS8.pdf`.

   The 2017 video selection is:

   > Overview of graphical models, loading Tetrad, Causal graphs and interventions
   > Hour 1 of the *2016 Summer Short Course on Causal Discovery*
   > Richard Scheines, Carnegie Mellon University
   > https://youtu.be/9yEYZURoE3Y

5. **(530/730) WEKA.**

   CIS 530 students: Read http://www.emunix.emich.edu/~sverdlik/DM/Version1a.html and make sure you understand the version space (candidate elimination algorithm) trace. Using WEKA v3.9 from **http://bit.ly/weka-download** (or the stable book version, v3.8.3, will also suffice), which you downloaded and installed for MP7, compare decision tree induction (J48), which you ran in MP7, on this Japanese-Economy-Car data set, in the Classify tab**.** You will need to create an ARFF file of this data set by yourself. Consult the WEKA documentation to learn how to do so.

   CIS 730 students: Do the above and also trace through the performance of Naïve Bayes.

   Turn in your solution as `PS8-5.pdf` or as part of `PS8.pdf`.

Class Participation (required)

Post your interim project report to the Interim Report thread on Canvas.

Also post to the PS8 thread any questions about any part of this assignment, about probabilistic reasoning and machine learning, or about your term project.

Coming Up Next

Machine Problem 9 (due Fri 16 Nov 2018) – Perception and Understanding, Part I: Artificial Neural Networks (ANNs) and Genetic and Evolutionary Computation (GEC). You will apply ANN and GEC to a pattern recognition or obstacle avoidance task for a simple mobile robot.

Problem Set 10 (due Fri 30 Nov 2018) – Perception and Understanding, Part II: Natural Language Processing (NLP), and Vision. You will solve problems and answer some discussion questions about perception and understanding. (This will include some practice final exam questions.)