

CSC421 Assignment 3. Summer 2018 (10pts)

Misunderstanding of probability may be the greatest of all impediments to scientific literacy.

Gould, Stephen Jay

Student Name:

Student Number:

Instructor George Tzanetakis

Question	Value	Mark
1	4	
2	3	
3	3	
Total	10	

1 Overview

So far in the assignments we have been mostly building things from “scratch” and not using existing libraries and frameworks as much. In this assignment you are asked to use existing tools to explore some of the ideas we have learned in class. I suggest software in Python for this purpose but if you can find similar software in other languages that is ok. The topics covered are CSP, machine learning with Naive Bayes and Discrete Bayesian Networks.

Don’t hesitate to contact the instructor via email or utilize the chat room of the ConneX course website for any questions/clarifications you might need.

The questions are marked as (*) necessary to pass, (**) expected and (***) exceptional.

IMPORTANT: YOU SHOULD ONLY SUBMIT A SINGLE PDF FILE WITH YOUR REPORT THROUGH CONNEX. ANY OTHER FORMAT WILL NOT BE ACCEPTED. THE ANSWERS SHOULD BE LABELED BY THE CORRESPONDING QUESTIONS NUMBERS

2 Constraint-satisfaction problems (4pts)

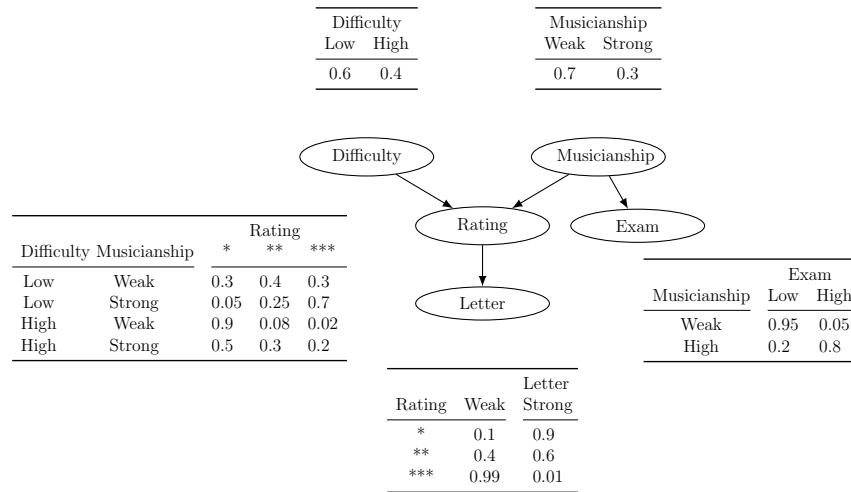
1. Using the conventions of the CSP library you are using solve the Cryptarithmic puzzle provided in Figure 5.2 of the AIMA textbook. **(3pts (ONLY UGRADS) - **)**
2. Similarly solve the problem of Waltz Filtering described in the following document <http://www.cs.cityu.edu.hk/~hwchun/research/PDF/PAJava99Waltz.pdf> using the library. There is a lot of information in the paper and it uses *JSolver* a java library for constraint programming. Your task is to transfer the constraints and problem described in the paper to *pgmpy*. This is easier than it might seem at first read but it does require some more effort than the simple example of the first question which is why this question is for graduate students or undergraduates who want an A+. **4pts (GRADS - NO NEED TO IMPLEMENT QUESTION 1), 1pt (UGRAD).**

3 Naive Bayes Text Classification using scikit-learn (3pts)

Using the movie review data from assignment 2 use the *scikit-learn* library to compute the classification accuracy using 10-fold cross-validation of the Naive Bayes Bernoulli and Naive Bayes Multinomial classifiers that are part of *scikit-learn*. Show your code and results. You will need to read the excellent documentation of *scikit-learn* for the details. **(3pts - ***)**

4 Discrete Bayesian Networks (3pts)

Consider how to compute the probability of a particular state for example the probability that the candidate has strong musicianship, the pieces played are not that difficult, the rating is 2 stars, the exam score is high, and the recommendation letter is weak. We can use the semantics of Bayesian networks



to compute this joint probability as follows:

$$\begin{aligned}
 &P(m = strong)P(d = low)P(r = ** | m = strong, d = low) \\
 &P(e = high | m = strong)P(letter = weak | **) \\
 &= 0.3 * 0.6 * 0.08 * 0.8 * 0.4 \\
 &= 0.004608
 \end{aligned}$$

For this question my suggestion is to use <https://github.com/pgmpy/pgmpy> which is a Python Library for Probabilistic Graphical Models for this assignment. Your deliverable should show all the code you used to solve the problem. Your task is to explore this particular discrete Bayesian Network.

1. Using *pgmpy* (or some other simialr library) show how the query provided as an example can be computed using exact inference (**1pt - ****)
2. Consider a particular musician, George and we want to know how likely he will get a strong recommendation letter. Knowing nothing else this probability is about 50.2%. If we find out that George is not a strong musician then that probability goes down to around 38.9%. Show how these numbers are derived using the software tool and exact inference (**1pt - ****)

3. Answer the previous two questions using some approximate inference algorithm. Comment on the difference **(1pt)**

Just to be clear you DO NOT need to do the exact or approximate inference by hand but just use the tool with the right conventions to get the results.

5 Grading

The submission should be a single PDF containing code snippets, text and figures with your answers in order. The questions are worth a total of 10 points. Grading will be based on the content of the answers as well as the quality of the report.