

CSC421 Assignment 4. Summer 2018 (10pts)

We must become more comfortable with probability and uncertainty.
Nate Silver

Student Name:

Student Number:

Instructor George Tzanetakis

Question	Value	Mark
1	4	
2	6	
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 Logic Programming and Hidden Markov Models.

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 Logic Programming (4pts)

The goal of this question is to contrast logic programming with traditional imperative programming. My suggestion is to use *logpy* <https://github.com/logpy/logpy> which is a logic programming library in Python. Alternatively you can use any logic programming language (SWI Prolog is the canonical example) and contrast it with an imperative or object oriented language.

Check the documentation of *logpy* for how to represent knowledge and form queries (the Simpsons example).

1. Using the same approach model the following knowledge: Darth Vader is the parent of Luke Skywalker and Leia Organa. Leia Organa and Han Solo are the parents of Kylo Ren. **(1 pt - *)**
2. Formulate the following queries: who is the parent of Luke Skywalker ? who are the children of Darth Vader ? **(1pt - *)**
3. Define the relationship grandparent and formulate the query who is the grandparent of Kylo Ren ? **(1pt - **)**
4. Write functions to answer the queries above using standard Python and not utilizing logic programming. Represent the family tree using standard data structures like dictionaries, tuples and lists. Write a few sentences contrasting the two approaches for this particular problem **(1pt - ***)**

3 Hidden Markov Models (6pts)

Consider using a markov modeling approach for a basketball player. The player can be in two states *healthy* or *injured*. Depending on the state the player chooses to shoot, dribble or pass with the following probabilities. Assume that the player always starts out healthy. From *healthy*, the transition probability to injured is 0.3; the probability of remaining *healthy* is 0.7. From injured, there is a 0.5 probability of staying injured and a 0.5 probability of changing back to healthy.

For this question I recommend using *hmmlearn*. <http://hmmlearn.readthedocs.io/en/latest/index.html>.

	Dribble	Pass	Shoot
Healthy	0.2	0.1	0.7
Injured	0.3	0.6	0.1

1. Write code that generates random sequences of *healthy*, and *injured* states based on the transition probability matrix provided above. **(1pt - *)**
2. Write code that generates random sequences of *dribble*, *pass*, *shoot* based on the markov model specified above. **(1pt - *)**
3. Use the fit method to learn a Hidden Markov Model for the data you generated. Use a sequence of 300 observations. Show the learned transition matrix and emission probabilities. How different are they from the original model used to generate the data ? **(2pts - **)**
4. Use the predict method to find the sequence of hidden states that generated your observations. Count how many errors the HMM model made by using the original states used for generation as ground truth **(2pts - ***)**.

4 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.