CSCI 3202: Introduction to Artificial Intelligence
Homework 7 (rev. 1): due 27. November 11:59pm to Moodle.

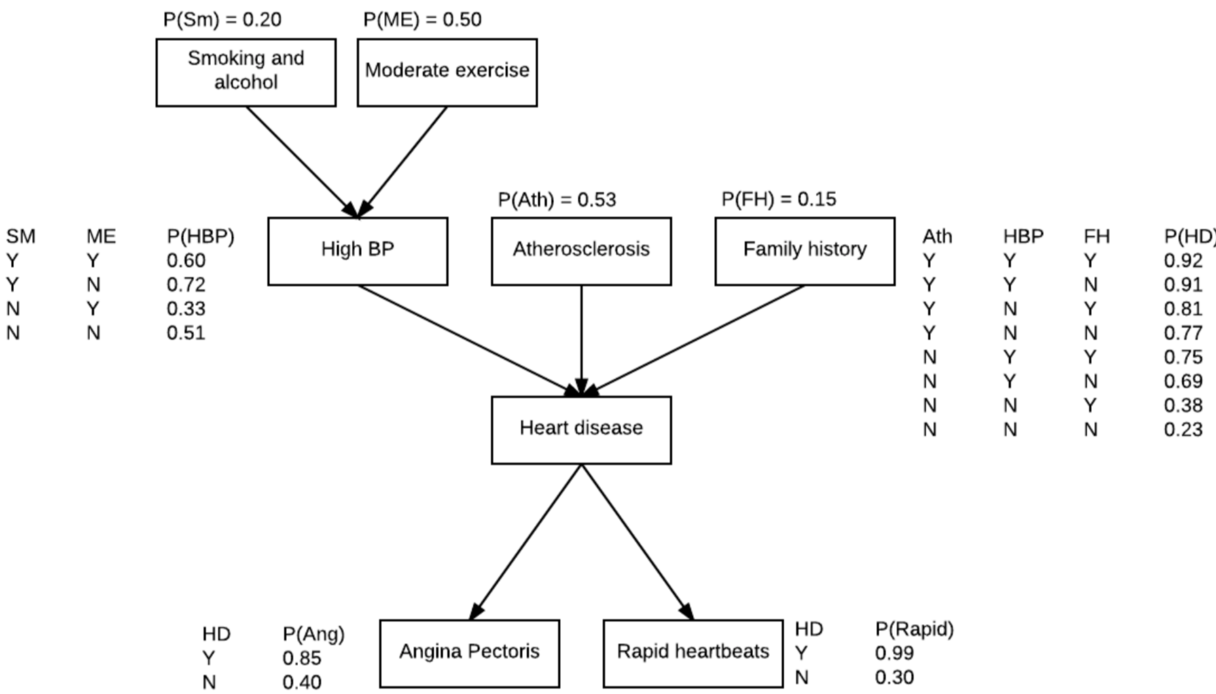
Release calendar for this homework:

+ 14. November: this prompt rev. 1 released with skeleton code.
- 18. November: CodeRunner quiz for Problem 1 available on Moodle.
- 21. November: CodeRunner quiz for Problem 2 available on Moodle.

Learning objective: Apply Bayesian Network models to probabilistic problems in multivariate distributions.

1 Bayesian network to model heart disease

The following Bayesian network is based loosely on a study that examined heart disease risk factors in 167 elderly individuals in South Carolina. Note that this figure uses Y and N to represent Yes and No, whereas in class we used the equivalent T and F to represent True and False Boolean values.



1.1 Creating your Bayes Net

Create a BayesNet object to model this. Below are the codes for the (conditional) probability `P` function and `BayesNode` class as well, that we used in class to represent the variable nodes and calculate probabilities. You can code this however you want, subject to the following constraints:

1. the nodes are represented using the `BayesNode` class and can work with the `P` function for probabilities,

2. your BayesNet structure keeps track of which nodes are in the Bayes net, as well as
3. which nodes are the parents/children of which other nodes.

Some *suggested* skeleton codes for a class structure are given. The suggestions for methods to implement are in view of the fact that we will need to calculate some probabilities, which is going to require us to `find_nodes` and `find_values` that nodes can take on.

Moodle Quiz Problem 1. Pass the Problem 1 unit tests (see skeleton code).

The tests may be run using:

```
tests_to_run = unittest.TestSuite()
tests_to_run.addTest(Tests_Problem1("test_onenode"))
tests_to_run.addTest(Tests_Problem1("test_twonode"))
unittest.TextTestRunner().run(tests_to_run)
```

1.2 Probability getters

Craft a function `get_prob(X, e, bn)` to return the normalized probability distribution of variable `X` in Bayes net `bn`, given the evidence `e`. That is, return $P(X|e)$. The arguments are:

- `X` is some representation of the variable you are querying the probability distribution of. Either a string (the variable name from the `BayesNode`) or a `BayesNode` object itself are good options.
- `e` is some representation of the evidence your probability is conditioned on. When given an empty argument (or `None`) for `e`, `get_prob` should return the marginal distribution $P(X)$.
- `bn` is your `BayesNet` object.

You may do this using the `enumeration` algorithm from class (pseudocode is in the book), or by brute force (i.e., use a few `for` loops). Either way, you should be using your `BayesNet` object to keep track of all the nodes and relationships between nodes so your `get_prob` function knows these things.

Moodle Quiz Problem 2. Use your `get_prob` function to calculate the following probabilities. Compare to the original Bayes net figure given to make sure the output passes these "unit tests."

1. The marginal probability of `Family History` is $P(FH = T) = 0.15P(FH = T) = 0.15$
2. The probability of not experiencing `Angina Pectoris`, given `Heart Disease` is observed, is $P(Ang = F|HD = T) = 1 - 0.85 = 0.15P(Ang = F|HD = T) = 1 - 0.85 = 0.15$
3. The probability of `High Blood Pressure`, given a person does `Smoke and/or use Alcohol` but does not get `Moderate Exercise`, is $P(HBP = T|Sm = T, ME = F) = 0.72P(HBP = T|Sm = T, ME = F) = 0.72$.

Moodle Quiz Problem 3. Use your `get_prob` function to calculate the following probabilities. These are more challenging so can't as easily be compared to the Bayes' Net diagram above.

1. The probability of an arbitrary individual having `Heart Disease`, $P(HD = T)$.
2. The probability that an individual does **not** have `Heart Disease`, given that `Rapid Heartbeat` was observed, $P(HD = F \mid Rapid = T)$
3. The probability of an individual having `High Blood Pressure` if they have `Heart Disease` and a `Family History`, $P(HBP = T \mid HD = T, FH = T)$
4. The probability that an individual is a `Smoker/Alcohol User` if they have `Heart Disease`, $P(Sm = T \mid HD = T)$
5. How would you expect the probability above to change if you also know the individual has `High Blood Pressure`? Verify your hypothesis by calculating the relevant probability.
6. How would you expect the probability above to change if you also know that the individual does **not** get `Moderate Exercise` (in addition to having `Heart Disease` and `High Blood Pressure`)? Verify your answer by calculating the relevant probability.

2 Bayesian network to model decision-making

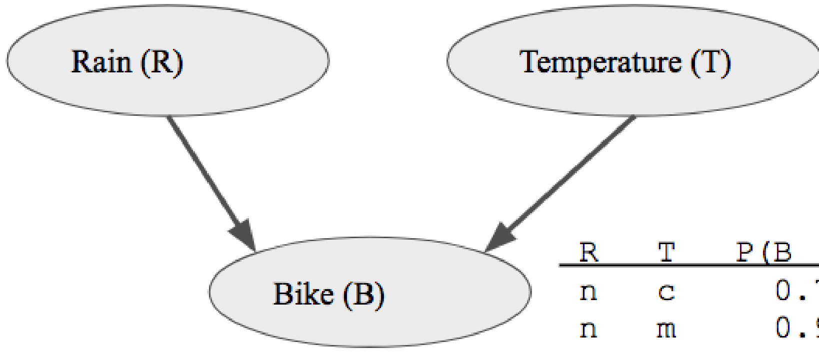
Let's consider using a Bayesian network to model our decision about whether or not to ride our bike to work today. This decision depends heavily on the weather, so let's focus on that.

In class, we focused on Boolean variables. For example, we might base our biking decision on whether or not it is raining. But in reality, it probably matters **how hard** it is raining. So suppose we break the variable `Rain` up into three discrete bins: `none`, `light` and `heavy`.

The temperature also factors into our decision. There is definitely a sweet spot, where temperatures are neither too warm nor too cold, so it is very likely we would enjoy riding our bike. So we can model the variable `Temperature` also using three discrete bins: `cold`, `moderate` and `warm`.

So a Bayesian network to model our decision for whether or not to bike to work could be as follows, where the first letter of each discrete bin is used to denote that variable value (i.e., `R=h` stands for heavy rain conditions).

R	P(R)
n	0.8
l	0.15
h	0.05



T	P(T)
c	0.3
m	0.6
w	0.1

R	T	P(B | R, T)
n	c	0.7
n	m	0.99
n	w	0.9
l	c	0.4
l	m	0.6
l	w	0.5
h	c	0.2
h	m	0.4
h	w	0.3

2.1 Setting up Bayes Networks to handle non-Booleans

In this order:

1. Modify the `P` probability function to be able to handle these ternary parent nodes.

2. Set up `BayesNode` objects for each of `Rain`, `Temp` and `Bike`, and create a `BayesNet` object to model the Bayesian network for this decision. Again, you can use whatever structure you wish for your `BayesNet`, but please use the `BayesNode` class. You may need to make minor modifications to the `BayesNode` class (e.g., changing/adding attributes), although none are strictly necessary.

> **Moodle Quiz Problem 4.** Verify that your modified probability function `P` is working by checking the following "unit tests" and giving their output. Compare to what you expect from the figure above.

1. The marginal probability of no rain is $P(Rain = n) = 0.8$

2. The marginal probability of light rain is $P(Rain = l) = 0.15$

3. The marginal probability of heavy rain is $P(Rain = h) = 0.05$

4. The probability of biking given that it is raining heavily and the temperature is cold, is $P(Bike = T \mid Rain = h, Temp = c) = 0.2$

2.2 Marginal and Conditional Probabilities

Make any necessary modifications to your `get_prob` function from Problem 1 so that you can use it to calculate marginal probabilities and conditional probabilities for this problem. It is possible that your function does not require any modifications.

Moodle Quiz Problem 5. Use `get_prob` to calculate the listed quantities below.

1. $P(Bike)$, the probability for whether or not you will ride your bike on any given day; and
2. the probability that you will ride your bike, given that it is lightly raining.

2.3 Inferring Temperature

We are trapped indoors because some jerk gave us a ton of Intro to Artificial Intelligence homework to do. Suppose we look out the window and see people biking. They sure do look like they're having fun! **Given** this information, we can actually make inferences regarding the temperature outside! What is the probability distribution for temperature, given that we observe people biking?

Moodle Quiz Problem 6. Compute this using your `get_prob` function for each of the possible temperatures.