

Assignment 5 Comprehension

B351 / Q351

Due: March 10th, 2020 @ 11:59PM

INSTRUCTIONS: Please submit your answers as ONE PDF file by the deadline in canvas. This is an individual assignment. Make sure you check your file to make sure you submitted what you intended to submit.

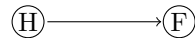
Here are the relevant sections in our book for this assignment:

Bayesian Network Section 7.4

Machine Learning Intro/Decision trees Chapter 8: Intro and Section 8.4

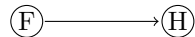
1 Comprehension Questions

1. Consider the following Bayesian network where H is the event of a person having a headache and F is the event of the person having the flu.



where $P(H) = 0.12$, $P(F | H) = 0.74$ and $P(F | \neg H) = 0.2$.

- (a) Compute the joint probabilities $P(H, F)$, $P(H, \neg F)$, $P(\neg H, F)$, and $P(\neg H, \neg F)$.
- (b) Determine the probabilities and conditional probabilities for the following Bayesian network:



so that it determines the same joint probabilities as in question 1(a). In other words, find $P(F)$, $P(H | F)$, and $P(H | \neg F)$ so that $P(H, F)$, $P(H, \neg F)$, $P(\neg H, F)$, and $P(\neg H, \neg F)$ are the same as in question 1(a).

- (c) Which Bayesian network makes more sense given the events H , and F , the one in 1(a) or 1(b)? Why?
2. Why do you separate training data from testing data in Machine Learning?

3. Saúl is trying to decide whether he should go out to eat or cook at home. He obtains a dataset of his teaching assistants and whether or not they went out to eat on various days. The attributes are the weather on the day ("Weather"), number of days since the last time they went to the grocery store ("Last Grocery Trip"), if they got paid that week ("Pay Day Week"), and if they have a car ("Car").

Name	Weather	Last Grocery Trip	Pay Day Week	Car	Result
Rowan	Rain	3-4 days	No	No	Stay In
Gabriel	Sun	3-4 days	No	Yes	Stay In
Andrew	Rain	≥ 5 days	Yes	Yes	Go Out
Niki	Snow	1-2 days	Yes	No	Stay In
Aaron	Sun	≥ 5 days	Yes	Yes	Go Out
Kyle	Snow	≥ 5 days	Yes	Yes	Stay In
David	Rain	3-4 days	No	Yes	Stay In
Jake	Sun	1-2 days	No	Yes	Stay In
Griffin	Snow	≥ 5 days	No	Yes	Go Out
Abe	Sun	1-2 days	Yes	Yes	Go Out

- (a) Calculate the entropy of the dataset (in other words, the entropy with respect to "Result").
 - (b) What's the entropy with respect to the attribute "Weather"?
 - (c) What is the information gain if the attribute "Last Grocery Trip" is split?
4. Explain the algorithm to build a decision tree.
 Note: There are a few ways to do this, but please explain the one talked about in class. More specifically, how do we decide which nodes to choose when we are building the tree? What's the priority?

2 Bonus

1. What is the smallest value that information gain can achieve for binary attributes? Prove your answer.