

# EECS 491 Assignment 2

Due Mon Mar 4 before midnight. 100 points total.

## Submitting assignments to Canvas

- For jupyter notebooks, submit the .ipynb file and a pdf or html export of the notebook. Make sure the output represents the latest state of your notebook. If you use interactive plots, make sure the output for the static file is representative of the points you wish to make. If you are not using notebooks, writeup your assignment using latex and submit a pdf with your code. The writeup should include relevant code with description if it can fit on a page. Do not include binaries or large data files.
- Use the following format for filenames:
  - EECS491-A2-yourcaseid.ipynb
  - EECS491-A2-yourcaseid.pdf
- If you have more than these two files, put all your files in a directory named EECS491-A2-yourcaseid . Then zip the directory and submit it with the name EECS491-A2-yourcaseid.zip . Do not use other compression formats.

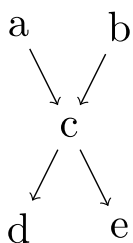
## Q1. Conditional Independence (10 pts)

Consider the following Bayes net.

In [1]:

```
1 using TikzGraphs, LightGraphs
2 g = DiGraph{5}
3 add_edge!(g, 1, 3); add_edge!(g, 2, 3); add_edge!(g, 3, 4); add_edge!(g, 3, 5)
4 TikzGraphs.plot(g, ["a", "b", "c", "d", "e"], options="font=\\large")
```

Out[1]:



1.1. (5 pts) Show that  $a$  is independent of  $b$  given no other information, i.e.

$$a \perp b \mid \emptyset$$

1.2. (5 pts) Prove or disprove the following using basic probability (i.e. not using d-separation)

$$a \perp b \mid e$$

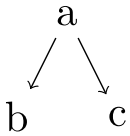
## Q2. Conditional Independence and Causality (5 pts)

Consider the following model

In [2]:

```
1 g = DiGraph(3)
2 add_edge!(g, 1, 2); add_edge!(g, 1, 3)
3 TikzGraphs.plot(g, ["a", "b", "c"], options="font=\\large")
```

Out[2]:



Show that this causal relationship suggested by the arrows does not necessarily hold, because the identical distribution can be represented by a model defined by different conditional distributions. What conditional independence assumption does this model make?

## Q3. Model Complexity, Free Parameters, and Simplifying Assumptions (20 pts)

3.1. (5 pts) Consider a general probability distribution with  $N$  variables  $x_1 \dots x_N$  each of which can have  $K$  values. What is the expression for the joint distribution in terms of conditional probabilities?

3.2. (5 pts) What is the total number of free-parameters required to specify this model? (Note: the term "free parameter" means a parameter that is unconstrained. For example a Bernoulli distribution to describe a coin flip has one free parameter  $\theta$  to describe, say, the probability of heads; the probability of tails must be  $1 - \theta$ , because the probability is constrained to sum to one.)

3.3. (5 pts) Now suppose that the complexity of the model is constrained, so that each variable depends on (at most)  $m$  other variables and is conditionally independent of the rest, i.e. each node has  $m$  parents and there are  $m$  root nodes. How many parameters are required to define this model?

3.4. (5 pts) Let us make one more simplifying assumption, which is that in addition to depending on only  $m$  variables, the conditional probability is described by a noisy-OR function (see Q3). What is the expression for the number of parameters in this case?

## Q4. Models of Conditional Probability (20 pts)

In Bayesian networks (or directed acyclic graphical models), the joint probability distribution is factored into the product of conditional probability distributions

$$p(x) = \prod_{i=1}^N p(x_i | \text{pa}(x_i))$$

As we used the previous problem, a simplifying assumption for the conditional probability is noisy-OR model

$$p(x_i | \text{pa}(x_i)) = 1 - (1 - \mu_{i0}) \prod_{j \in \text{pa}(x_i)} (1 - \mu_{ij})^{x_j}$$

where  $j$  is an index over the parents of  $x_i$ . Note that the exponent  $x_j$  is either 0 or 1 so the term is either 1 or  $1 - \mu_{ij}$  depending on the state of the parent  $x_j$ .

4.1 (5 pts) Show that the noisy-OR function can be interpreted as a "soft" (i.e. probabilistic) form of the logical OR function, i.e. the function gives  $x_i = 1$  whenever at least one of the parents is 1.

4.2 (5 pts) What is the interpretation of  $\mu_{i0}$ ? Provide a clear explanation.

Another choice for the conditional probability is a sigmoid function

$$p(x_i | \text{pa}(x_i)) = \sigma \left( w_{i0} + \sum_{j \in \text{pa}(x_i)} w_{ij} x_j \right), \quad \text{where } \sigma(a) = \frac{1}{1 + e^{-a}}$$

where  $\sigma(a)$  is the logistic sigmoid function.

4.3 (5 pts) Contrast the noisy-OR function and the sigmoid mathematically. Is one more general than the other? Can each compute unique functions?

4.4 (5 pts) Think of two examples, one for the noisy-OR and one for the sigmoid, that contrast the way these functions model the conditional dependencies. Explain how each is appropriately modeled by one function but not the other.

Q5. Car Troubles (Barber Exercise 3.6) (25 pts)

- 5.1 (10 pts) Calculate the  $p(f = \text{empty} | s = \text{no})$  , the probability of the fuel tank being empty given that the car does not start. Do this "by hand", i.e in manner similar to the **Inference** section in Barber 3.1.1. Use the probabilities given in the exercise. Show your work.
- 5.2 (5 pts) Implement this network using a toolbox for probabilistic models (e.g. `pgmpy` or `BayesNets.jl` ). Use this to verify that your derivation and calculations are correct for the previous problem.
- 5.3 (10 pts) Suppose you have loaned this car to a friend. They call call you and announce, "the car won't start". Illustrate your diagnostic and inference process by using the model to show how your beliefs change as you ask questions. Your friend can only tell you the states of  $t$  and  $g$  (and you already know  $s$ ). Use two different scenarios.

Exploration (20 pts)

Like in the first assignment, in this problems, you have more lattitude and are meant to do creative exploration. Like before you don't need to write a book chapter, but the intention is for you to go beyond what's been covered above.

Implement a belief network of your own choosing or design. It should be more complex that the examples above. It should be discrete (we will cover continous models later). Use the model to illustrate deductive inference problems.

The grading is necessarily subjective, but here is a rubric:

- Was scenario you are trying to model clearly described? (5 pts)
- Were the variables and states well-chosen and clearly explained? (5 pts)
- How well did the examples to illustrate the model? (5 pts)
- Did the model go beyond or is distinct from what was already convered in the questions above? For example, larger numbers of nodes with more complex interactions or more variable states or using simplifying assumptions for conditional probabilities like noisy-OR, etc. (5 pts)

In [ ]:

1	
---	--