Probabilistic Models

A probabilistic model is a joint distribution over a set of variables:

$$P(X_1, X_2, \dots, X_n)$$

We can represent this quantity through the Chain Rule:

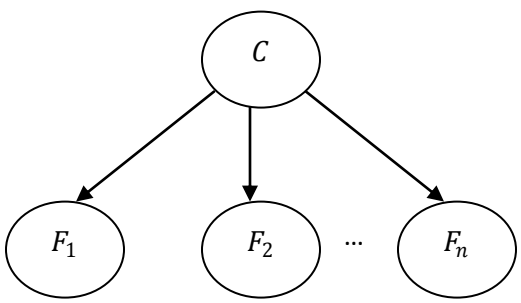
$$P(X_1, X_2, \dots, X_n) = \prod_i P(X_i | X_{i+1}, \dots X_n)$$

Using a Bayes' net, we can reduce this redundancy to:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} P(X_i \,|\pi(X_i))$$

Again, we define $\pi(X_i)$ equal to the parents of X_i.

Naïve Bayes Classifier



Using the chain rule:

$$\begin{aligned}
&p(C, F_1, \dots, F_n) \\
&\quad = p(C)\ p(F_1, \dots, F_n | C) \\
&\quad = p(C)\ p(F_1|C)\ p(F_2, \dots, F_n | C, F_1) \\
&\quad = p(C)\ p(F_1|C)\ p(F_2|C, F_1)\ p(F_3, \dots, F_n | C, F_1, F_2) \\
&\quad = p(C)\ p(F_1|C)\ p(F_2|C, F_1)\ p(F_3|C, F_1, F_2)\ p(F_4, \dots, F_n | C, F_1, F_2, F_3)
\end{aligned}$$

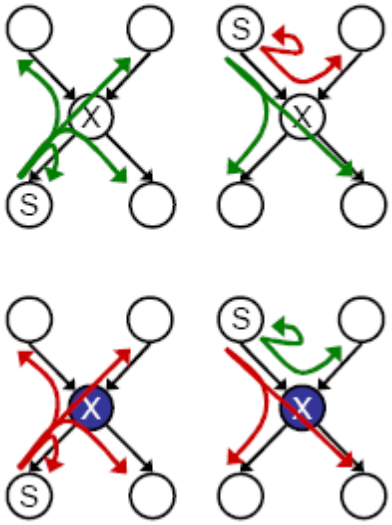
Using conditional independence for Bayes' Nets:

$$P(C, F_1, \dots, F_n) = P(C) * \prod_i^n P(F_i|C)$$

Conditional Independence in Bayes' Net

Reachability (the Bayes' Ball)

- Correct algorithm:
  - Shade in evidence
  - Start at source node
  - Try to reach target by search
  - States: pair of (node X, previous state S)
  - Successor function:
    - X unobserved:
      - To any child
      - To any parent if coming from a child
    - X observed:
      - From parent to parent
  - If you can't reach a node, it's conditionally independent of the start node given evidence



Future topics:

- Sampling and Variable Elimination
- Decision trees
- More prediction/learning with HMMs

Question 1 (Directed graphical models and conditional probabilities)

A standard test for prostate cancer, the prostate specific antigen (PSA) test, has a sensitivity of around 95% and a specificity of around 55%. (The sensitivity is the conditional probability of a positive test result given that the disease is present; the specificity is the conditional probability of a negative test result given that the disease is absent.) The probability of developing prostate cancer is age-dependent.

Suppose that:

- The probability that a sixty-year-old patient has prostate cancer is 2%, and this probability increases with age.
- The probability that a sixty-year-old patient has a more benign prostate disease, BPH, is 8%, and this probability increases with age.
- Given the patient's age the event that he has BPH is conditionally independent of the event that he has prostate cancer.
- Age and presence or absence of BPH does not affect the sensitivity or specificity of the PSA test.
- Independent of age, the probability of an enlarged prostate is 10% for healthy patients, 85% for patients with BPH, 60% for patients with prostate cancer, and 85% for patients with both diseases.

(a) Draw the graph of a directed graphical model that describes the joint probability distribution. List the constraints on entries in the conditional probability tables that have been specified in the information presented above.

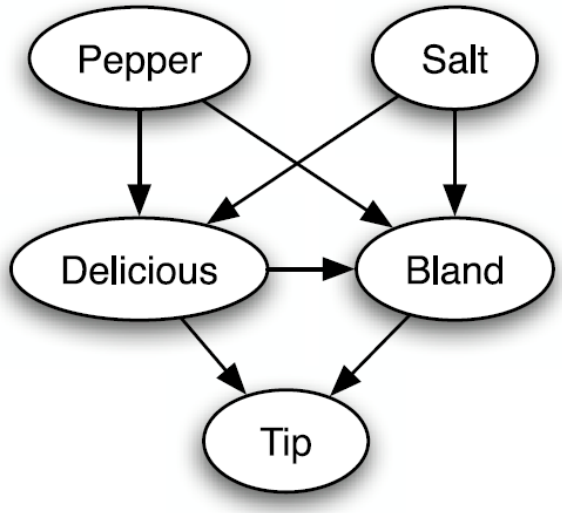
(b) If a sixty-year-old patient tests positive on the PSA test, what is the probability that he has prostate cancer?

(c) If the patient also has an enlarged prostate, what is the probability that he has prostate cancer?

(d) If a sixty-year-old patient has an enlarged prostate, what is the probability that he has BPH?

Question 2

An ambitious Bongo Burger employee wants to maximize tips, which he thinks depend on the quality of the chef's cooking. When the chef uses more Pepper and more Salt, more people say "Delicious" and fewer people say "how Bland". Better comments come with better tips. He draws the following Bayes' net with 5 Boolean variables to model the scenario.



a. Which of the following are true of this network? Note that a marginal distribution over a subset of variables in a Bayes' net is the result of summing out all other variables from the full joint distribution.

(Circle whichever are true, 1 pts each)

i. It can represent any joint probability distribution over Salt, Pepper, Delicious, Bland, Tip.

ii. For any marginal distribution over Salt, Delicious, Bland, the network can represent at least one joint distribution with that marginal distribution.

iii. For any marginal distribution over Salt, Tip, the network can represent at least one joint distribution with that marginal distribution.

b. How many total probability entries are there in the conditional probability tables for this network. Note: not the number of degrees of freedom, which may be fewer; a coin flip has two probabilities, but only one degree of freedom, and we're asking for the larger number.

c. List all of the marginal independence assertions made by this graph about pairs of variables.

d. List all of the conditional independence assertions made by this graph about pairs of variables.

e. For one of the conditional independence assertions above, briefly explain how you might test its correctness from a set of observations.

f. "No, no, no," the chef replies, "People tip because the food is Tasty (influenced by salt and pepper), but comments are more complicated. Customers say Delicious or Bland in part based on whether they like the food, but mostly based on how Polite they are." Draw a new Bayes' net appropriate to this scenario, which includes variables for Tasty and Polite, but which has as few arcs as possible. You do not need to specify the CPTs.

g. Given $P(Pepper = True) = P(Salt = True) = 0.5$, the CPT below and the Bayes' net you just constructed for part (e), can you give $P(Pepper = True, Salt = False | Tasty = False)$? If so, compute it. If not, explain why not.

Pepper	Salt	P(Tasty=True|pepper, salt)
True	True	0.8
True	False	0.6
False	True	0.6
False	False	0.1

h. What is the (marginal) probability of the food being Tasty?

i. How might the chef alter her seasoning habits to increase $P(Tasty)$ without using any more salt or pepper overall? What would $P(Tasty)$ be after this change? Would this change alter the Bayes' net you drew for (e)? If so, how? If not, why not?