CSC242: Introduction to Artificial Intelligence
Homework 4: Probability and Introduction to Machine Learning

1. True or false: A prior (or unconditional) probability takes into account all relevant other information

2. Posterior (conditional) probabilities are conditional on what?

3. Prove that $P(a|b \wedge a) = 1$. Use the definition of conditional probability and some basic properties of conjunction.

4. Suppose we have the following random variables and domains (values) to describe the state of a car at a repair shop:

   - $\mathit{Problem} : \{\mathit{brakes}, \mathit{clutch}, \mathit{electrical}, \mathit{steering}, \mathit{tires}\}$
   - $\mathit{Rattling} : \{\mathit{true}, \mathit{false}\}$
   - $\mathit{Squeaky} : \{\mathit{true}, \mathit{false}\}$
   - $\mathit{Type} : \{\mathit{economy}, \mathit{midrange}, \mathit{luxury}\}$

   Give an English translation of the following conditional probability statement using these variables and values and the notational conventions in the book:

   $$P(\mathit{clutch}|\mathit{rattling} \wedge \neg \mathit{squeaky} \wedge \mathit{luxury}) = 0.4$$

5. Assume a random variable *Cuisine* with values $\{\mathit{american}, \mathit{japanese}, \mathit{chinese}, \mathit{french}, \mathit{polish}\}$ representing the type of meal served in a cafeteria for lunch. What does the following probability statement say?

   $$\mathbf{P}(\mathit{Cuisine}) = \langle 0.5, 0.2, 0.2, 0.1, 0 \rangle$$

6. Using the random variables from the car repair shop example above, draw (or describe) tables showing the elements of the following joint probability distributions (of course you can't fill in the values):

   (a) $\mathbf{P}(\mathit{Problem}, \mathit{Type})$

(b) $\mathbf{P}(\mathit{Problem}, \mathit{Type}, \mathit{Rattling})$

(c) $\mathbf{P}(\mathit{Problem}, \mathit{Type}, \mathit{Rattling}, \mathit{Sqeaky})$

7. What does it mean for two random variables to be independent? Be specific and use formal notation where possible.

8. What does it mean for two random variables to be conditionally independent? Be specific and use formal notation where possible.

9. Why are independence assertions useful for inference?

10. State Bayes' Rule briefly and formally.

11. What are the components of a Bayesian Network for a set of random variables $\{X_i\}$?

12. Exact inference in Bayesian Networks relies on two properties:

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

and

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \mathit{parents}(X_i)).$$

Briefly explain what these equations mean and how they are combined in inference.

13. Is exact inference for Bayesian networks feasible? Why or why not?

14. What does it mean for a probability estimate to be *consistent*. Be brief.

15. What are the main strength and main weakness of rejection sampling?

16. How does likelihood weighting improve on rejection sampling?

17. Briefly explain how Gibbs sampling works.

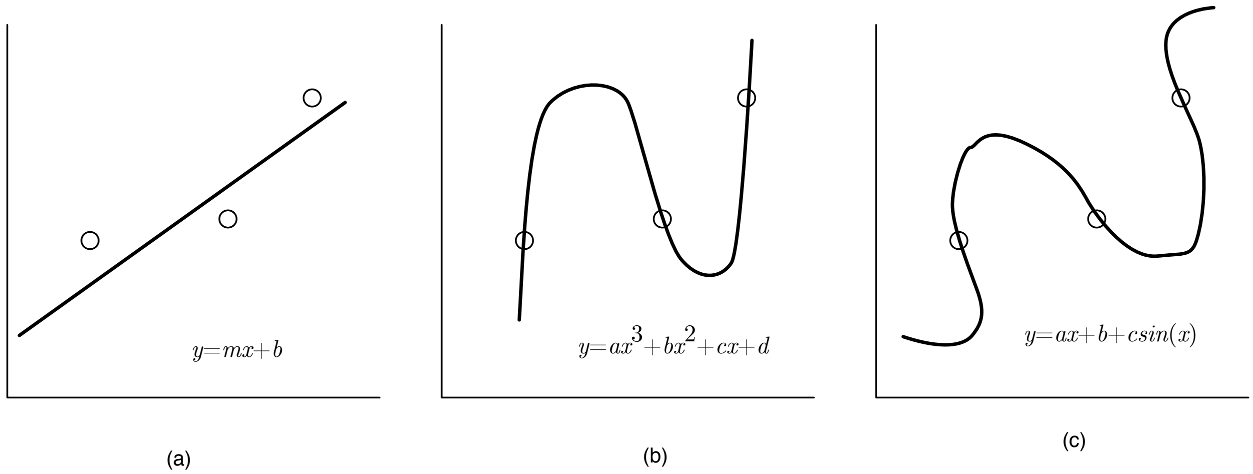
18. For each of the following types of learning, briefly describe what is learned and what knowledge, data, or feedback the learning agent receives to help it learn.

    (a) Unsupervised learning

    (b) Supervised learning

    (c) Reinforcement learning

19. The figure below shows a data set fit with several different function hypotheses.



Briefly compare the different hypotheses according to each of the following criteria:

(a) Simplicity of the hypothesis

(b) Goodness of fit

(c) Generalization

20. Why (besides the philosophical or aesthetic appeal of Occam's Razor) do we prefer simpler hypotheses for machine learning?

21. Write down the Propositional Logic sentence that is equivalent to the decision tree shown in Figure 18.6 (the one induced from the examples by the Decision-Tree-Learning algorithm).

22. When learning a decision tree, suppose that you have considered all the attributes but still you are left with a subset of the examples that cannot be classified uniquely. What does this mean? Could it happen in real life? Why or why not?

23. Explain briefly how to get the most out of your data when evaluating it using the training set/testing set paradigm.