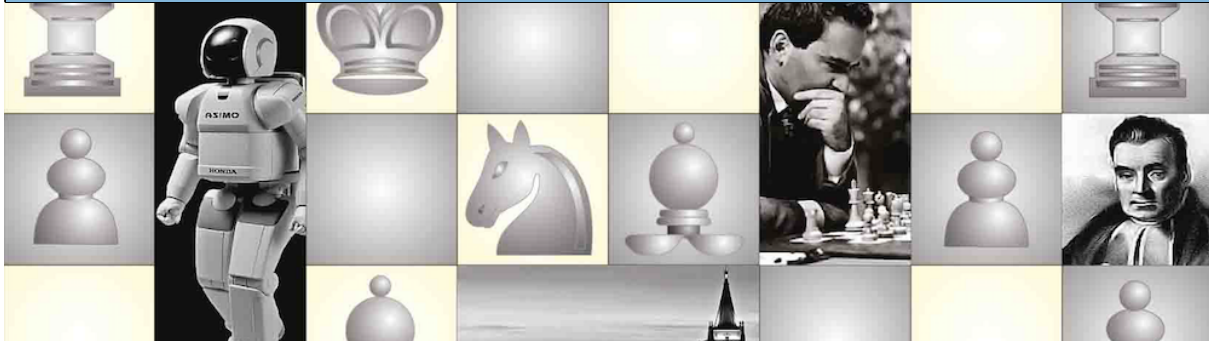


Fall 2024 - CSC545/645 Artificial Intelligence - Assignment 8



Due date: Tuesday, November 5, 2024, 2:00 pm. Please create a folder called `assignment8` in your local working copy of the repository and place all files and folders necessary for the assignment in this folder. Once done with the assignment, add the files and folders to the repo with `svn add files, folders` and then commit with `svn ci -m "SOME USEFUL MESSAGE"` files, folders.

Exercise 8.1 [20 points]

Read chapter "Probabilistic Reasoning" of the textbook.

1. You are a witness of a night-time hit-and-run accident involving a taxi in Athens. All taxis in Athens are blue or green. You swear, under oath, that the taxi was blue. Extensive testing shows that under dim lighting conditions, discrimination between blue and green is 75% reliable.
 - (a) Is it possible to calculate the most likely color for the taxi? (Hint: distinguish carefully between the proposition that the taxi *is* blue and the proposition that it *appears* blue.)
 - (b) What now, given that 9 out of 10 Athenian taxis are green? (Adapted from [Pea88])

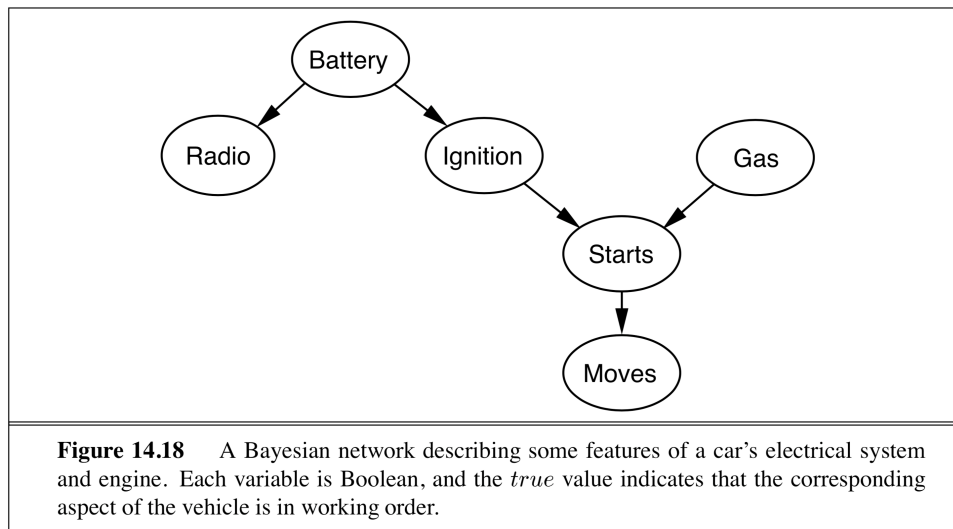
[4 points]

2. Text categorization is the task of assigning a given document to one of a fixed set of categories, based on the text it contains. Naive Bayes models are often used for this task. In these models, the query variable is the document category and the "effect" variables are the presence or absence of each word in the language; the assumption is that words occur independently in documents, with frequencies determined by the document category.
 - (a) Explain precisely how such a model can be constructed, given as "training data" a set of documents that have been assigned to categories.
 - (b) Explain precisely how to categorize a new document.
 - (c) Is the independence assumption reasonable? Discuss.

[6 points]

3. Consider the network for car diagnosis shown in figure below.
- (a) Extend the network with the Boolean variables *IcyWeather* and *StarterMotor*.
 - (b) Give reasonable conditional probability tables for all the nodes.
 - (c) How many independent values are contained in the joint probability distribution for eight Boolean nodes, assuming no conditional independence relations hold among them?
 - (d) How many independent probability values do your network tables contain?
 - (e) The conditional distribution for *Starts* could be described as a **noisy-AND** distribution. Define this family in general and relate it to the noisy-OR distribution.

[10 points]



References

- [Pea88] Judea Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Elsevier, 1988.