

COMP9016 Assignment #2

Dr Ruairi O'Reilly

November 10, 2020

1 KNOWLEDGE REPRESENTATION

The purpose of this assignment is to assess the following **Learning Outcomes**:

- LO1 Appraise domain specific formalisms used in knowledge representation schemes.
- LO2 Compare and contrast current knowledge representation approaches integrated in systems relevant to AI.
- LO3 Select, apply and evaluate a knowledge representation scheme for a specified domain.
- LO4 Design and implement KR formalisms for a real-world data set.
- LO5 Interpret, critique and communicate the suitability of data visualisation techniques used in conjunction with the design of KR formalisms and the analysis of the resulting output.

To that end, you are going to build upon the theory discussed in lectures and the practical work carried out as part the course.

1.1 PROBABILITY

1.1.1 PROBABILITY DISTRIBUTION TABLE

You have been studying hard on your latest assignment and asked your fellow students for some tips:

- Tip 1 “Study hard and you will do well, fail to do so and you will not”

- Tip 2 “Get plenty of rest and you will do well, fail to do so and you will not”
- Tip 3 “Set an an alarm and you will get up in time, fail to do so and you will not”.

You ask your fellow classmates which tips they adhere to and how frequently. After recording their responses you derive the following table:

	Never	Rarely	Sometimes	Often	Always
Tip 1	1	4	6	12	23
Tip 2	12	4	12	4	2
Tip 3	24	2	5	4	4

Create a probability distribution table using appropriate variable names.

1.1.2 BAYESIAN NETWORKS

You have been given the following Random Variables:

- Fossil Fuels
- AI
- GlobalWarming
- Renewable Energy
- Traffic
- Employed

Construct a model of the world using these variables (Note: this is subjective but there should be a rationale behind your world view and your thought process articulated). Implement a Bayesian Network.

- Provide a visual depiction of the network
- Detail the associated Conditional Probability Tables.
- Demonstrate querying the network.

Write a clear and concise description regarding Bayesian Networks, the utility of such and the associated pro's and cons. Support this with examples where possible.

1.2 LEARNING: DEVELOPING AND EVALUATING A MODEL

“Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.” You will be implementing a Naive Bayes classifier and discussing the theory behind its operation from the perspective of probability, conditional independence, and Bayes theorem.

1.2.1 DATA

Choose an appropriate multivariate datasets with a limited number of classes from: <https://archive.ics.uci.edu/ml/index.php> Take a subset of data from the dataset:

- Compute the “Prior” probabilities for each of the classes.
- Compute the probability of evidence.
- Compute the probability of likelihood of evidences (numerator).

Articulate your understanding of each as applied to Bayesian networks. Tip: Demonstrate the suitability of a dataset or highlight why it is not suitable. Use appropriate visualisations to aid your rationale.

Write a clear and concise description regarding the dataset and its appropriateness for a Naive Bayes Classifier and probabilistic inference in general.

1.2.2 NAIVE BAYES LEARNER

Based on your studying and reviewing of “learning.ipynb” implement the Naive Bayes learner and evaluate its performance for the selected dataset. Choose appropriate methodologies and/or variants in its utilisation. Present appropriate visualisations demonstrating its performance. Write a clear and concise description of the Naive Bayes Classifier considering the headings outlined: probability, conditional independence, and Bayes theorem

2 TIPS

- **Review** “probability.ipynb”, “learning.ipynb” and “learning_apps.ipynb”.
- **Tip for all Parts:** Do not provide masses of text lifted from the text/online or wikipedia. Be your own editor, focus on what is important and articulate that in a rationale way that is technically accurate. Demonstrating a capacity to explain complex topics in a clear and concise manner is indicative of a higher level of understanding and will be rewarded with an appropriate grade. Use results derived from experimental runs to back your claims and make sure your narrative is coherent and consistent.

3 SUBMISSION

You submission will contain:

- A 1500 word report ≤ 7 pages in “.pdf” format.
- A **single** “.py” file

Prepare your submission by archiving it as a single “.zip” file using the following naming convention.

“A2_COMP9016_<Surname>_<First name>_<Student Number>.zip”

e.g.

“A2_COMP9016_OReilly_Ruairi_R123456.zip”

3.1 WORD COUNT - 1500 WORD MAX

There are three distinct questions as part of this assignment Q1.1-Q1.3. In answering these it is important that you demonstrate an understanding of the underlying theory as applied in the context of your environment. There is a suggested word limit of 500 words per question, your ability to articulate your answer succinctly is indicative of a deeper understanding in itself. I **highly recommend** the use of well formatted tables for any experimental results and/or diagrams/visualisations if they aid the message you are trying to convey. Boilerplate text should be kept succinct and to the point.

Prepare your solution by the due date as a single “.pdf” file using the following naming convention.

“A2_COMP9016_<Surname>_<First name>_<Student Number>.pdf”

e.g.

“A2_COMP9016_OReilly_Ruairi_R123456.pdf”

3.2 CODE SUBMISSION

A **single** “.py” file should be prepared containing all code for questions 1.1-1.3. Encapsulate these in three individual functions called from the end of your submitted “.py” file. **NB:** Your code will be run from a “subdir” of the AIMA libraries (I have already provided a solution for importing these libraries from a parent directory - [see link](#) - I expect you to follow these guidelines).

Prepare your solution by the due date as a single “.py” file using the following naming convention.

“A2_COMP9016_<Surname>_<First name>_<Student Number>.py”

e.g.

“A2_COMP9016_OReilly_Ruairi_R123456.py”

3.3 SUBMISSION PENALTIES

- Late submissions will be penalised (-10 **points** for ≤ 1 week late, -20 **points** for > 1 week late AND ≤ 2 weeks late and **no submission accepted** after 2 weeks.
- Specified deadline is a **hard deadline** - requests for extension will not be considered (there is a formal process for individual extenuating circumstances that can be applied for via your programme coordinator e.g. medically certified illness, bereavement etc.).
- Failure to comply with submission guidelines will also result in a penalty (-10 **points**).

3.4 SUBMISSION DATE

This assignment is due on Sunday of Week 10 by 23:59.