CIT, DEPARTMENT OF COMPUTING

COMP9016 Assignment #2

Dr Ruairi O'Reilly

December 17, 2019

0.0.1 KNOWLEDGE REPRESENTATION

Design, develop and deploy a KR solution for a real world problem domain. e.g. a system for automating the interpretation of biomedical data requires a KR scheme that enables the translation of data such that the integrity of all domain specific information is maintained. Provide a rationale for the chosen visualisation approach taken at both the design and analysis stage. To this end, you are going to combine probabilistic uncertainty and learning in the development and evaluation of an model for KR formalisms.

1 Q1 - BASICS

Q1 is a basic evaluation of your coding ability.

1.1 PROBABILITY DISTRIBUTION TABLE - 5 MARKS

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 give your classmates to see which tips they adhere to and how often, after being surveyed they come back with the following:

	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.

1.2 Bayesian Networks - 20 marks

You have been given the following Random Variables:

- Fossil Fuels
- AI
- Global Warming
- 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 using the BayesNet class from the AIMA.

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

2 Q2 - 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.

2.1 DATA - 15 MARKS

Choose two appropriate multivariate datasets with a limited number of classes from: https://archive.ics.uci.edu/ml/index.php
Take a subset of data from and for each 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.

2.2 NAIVE BAYES LEARNER - 20 MARKS

Based on your studying and reviewing of "learning.ipynb" implement the naive bayes learner and evaluate its performance for the selected datasets.
Choose appropriate methodologies and/or variants in its utilisation.
Present appropriate visualisations demonstrating its performance.

3 TIPS

- **Tip for Part 1**: Review "probability.ipynb"
- **Tip for Part 2**: Review "learning.ipynb"
- **Tip for Part 2**: Review "learning_apps.ipynb"

4 SUBMISSION DATE & FORMAT

- This assignment is due by 23:59 at the end of Week 14 (Sun Dec 22nd)
- Your submission should include code (either a single python (.py) or jupyter notebook file (.ipynb))
- Your submission should include a report approx 1500 - 2000 words. (This can be a pdf or integrated in the jupyter notebook)
- Your report **should not** simply restate verbatim text taken from online sources/reference material. It should articulate your understanding of the underlying theory.