**CPSC 583 Fall 2019**

**Homework #3**

**Due 12/12/2019 by 11:55pm on Titanium**

**You may work in groups of two. Only one person needs to turn in the assignment on Titanium. Upload all your answers as a single PDF file**.

**1.** Consider the following table in a relational database storing the assignment of courses to classrooms in a university.

|  |  |  |
| --- | --- | --- |
| Course number (primary key) | Room | Department |
| CPSC-583 | CS-110B | ComputerScience |
| CPSC-597 | CS-110B | ComputerScience |
| CPSC-473 | CS-406 | ComputerScience |

Convert the relational data into a Semantic Web representation. Use the following two properties:

http://example.org/is-located-in

http://example.org/is-offered-by

1. Give the **triple** representation of the Semantic Web data. The first triple is already given (ignoring the full URI for subject and object).

PREFIX ex: http://example.org/

CPSC-583 ex:is-located-in CS-110B .

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Give the **graph** representation of the Semantic Web data. The first edge is already given.

is-located-in

**2**. Read the attached paper “Understanding and Adopting Semantic Web Technology” by John Kuriakose (Cutter IT Journal, 22(9), pages 10-18). Based on this paper, and optionally your own experiences, write a short essay (approximately 500 words) on what are the opportunities and challenges for enterprise-scale businesses when adopting Semantic Web technology. (Your answer must be typed.)

**3**. The SPARQL queries that were demonstrated in class are in the attached sparql\_examples.txt file. These queries were executed on the SPARQL endpoint for the dbpedia dataset: <http://dbpedia.org/snorql/>. Write additional queries to list:

1. all the systems of government in India (use property dbo:governmentType)
2. all countries (names in English only) which are republics (i.e., dbo:governmentType is :Republic)
3. all countries (names in English only) which are republics *and* have a parliamentary system (i.e., dbo:governmentType is :Republic and also :Parliamentary\_system)

**4.** Consider the following dataset with weather conditions on ten winter days (Cloudy, Freezing) and if an accident occurred on that day.

|  |  |  |  |
| --- | --- | --- | --- |
| Day | Cloudy | Freezing | Accident? |
| 1 | *False* | *True* | *No* |
| 2 | *True* | *False* | *Yes* |
| 3 | *True* | *False* | *No* |
| 4 | *True* | *True* | *Yes* |
| 5 | *True* | *False* | *Yes* |
| 6 | *True* | *False* | *No* |
| 7 | *True* | *True* | *No* |
| 8 | *True* | *True* | *Yes* |
| 9 | *True* | *False* | *No* |
| 10 | *False* | *True* | *Yes* |

1. Apply the ID3 decision tree learning algorithm to build a complete Decision Tree to decide if an accident is likely to occur on a given winter day. **Show the decision tree and calculations**. [To make calculations easier, you can use Table 1.]
2. For your decision tree, what is the accuracy on the given dataset?
3. Does there exist a decision tree (that only uses Cloudy and Freezing as input for its decision) that gives 100% accuracy on the given training data? Justify your answer (1-2 sentences).

**5.** For the same data in Problem 4:

1. Construct a **naïve Bayes classifier** to predict if an accident will occur on a particular day.

Prior probabilities:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Conditional probabilities:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Using the above naïve Bayes classifier, determine if an accident is likely to occur on a day that is both Freezing and Cloudy. Show calculations.
2. What is the accuracy of your naïve Bayes classifier on the given dataset?

**Optional (*No* extra credit):**

You can verify your answers to Q4 and Q5 by using the Weka machine learning software: <http://www.cs.waikato.ac.nz/~ml/weka/>

Weka includes both the ID3 (may need to be added using Weka’s built-in package manager) and the naïve Bayes classifiers. The training data is provided as a CSV file for import into Weka.

Table 1: Table of entropies for all combinations of n, p=1,2,...,10

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | n=1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| p=1 | 1.000 | 0.918 | 0.811 | 0.722 | 0.650 | 0.592 | 0.544 | 0.503 | 0.469 | 0.439 |
| 2 | 0.918 | 1.000 | 0.971 | 0.918 | 0.863 | 0.811 | 0.764 | 0.722 | 0.684 | 0.650 |
| 3 | 0.811 | 0.971 | 1.000 | 0.985 | 0.954 | 0.918 | 0.881 | 0.845 | 0.811 | 0.779 |
| 4 | 0.722 | 0.918 | 0.985 | 1.000 | 0.991 | 0.971 | 0.946 | 0.918 | 0.890 | 0.863 |
| 5 | 0.650 | 0.863 | 0.954 | 0.991 | 1.000 | 0.994 | 0.980 | 0.961 | 0.940 | 0.918 |
| 6 | 0.592 | 0.811 | 0.918 | 0.971 | 0.994 | 1.000 | 0.996 | 0.985 | 0.971 | 0.954 |
| 7 | 0.544 | 0.764 | 0.881 | 0.946 | 0.980 | 0.996 | 1.000 | 0.997 | 0.989 | 0.977 |
| 8 | 0.503 | 0.722 | 0.845 | 0.918 | 0.961 | 0.985 | 0.997 | 1.000 | 0.998 | 0.991 |
| 9 | 0.469 | 0.684 | 0.811 | 0.890 | 0.940 | 0.971 | 0.989 | 0.998 | 1.000 | 0.998 |
| 10 | 0.439 | 0.650 | 0.779 | 0.863 | 0.918 | 0.954 | 0.977 | 0.991 | 0.998 | 1.000 |