**CS 69995 & CS 79995 ST: Probabilistic Data Management**

**Homework 1**

**Instructor:** Xiang Lian

**Due Date:** Please refer to the course website

1. Please list 3 real applications that involve imprecise and uncertain data. Explain how the uncertainty (or probability) can be obtained for probabilistic data. **[30 points]**

2. What are the categories of uncertain data? Please explain the criterion you classify uncertain data. **[20 points]**

3. In Figure 1, assume that uncertain object *a* has 6 possible instances, and each of the rest uncertain objects, *b* ~ *e*, has 2 possible instances. Instances from the same uncertain object have equal appearance probabilities. **[20 points]**

3(a). How many possible combinations of object instances in this uncertain database that can appear in the real world? [10 points]

3(b). What is the *nearest neighbor* (NN) probability of uncertain object *d* when *a* is located at the red point? [10 points]



**Figure 1.** An example of probabilistic database.

4. In Table 1, please answer the following questions: **[30 points]**

4(a). How many possible worlds are there in the probabilistic database in Table 1? [10 points]

4(b). Please list all possible worlds and their appearance probabilities of the probabilistic database in Table 1. [20 points]

|  |  |  |  |
| --- | --- | --- | --- |
| **x-tuple** ***ti*** | **Alternative *tij*** | **Attribute** | **Probability** |
| *t*1 | *t*11 | 4 | 0.3 |
| *t*12 | 5 | 0.4 |
| *t*2 | *t*21 | 2 | 0.7 |
| *t*3 | *t*31 | 11 | 0.6 |
| *t*32 | 15 | 0.3 |
| *t*33 | 12 | 0.1 |

**Table 1.** An example of probabilistic table.

**Solution:** (You can add more rows in the table for possible worlds below)

|  |  |
| --- | --- |
| **Possible World, *PWi*** | **Appearance Probability, *Pr*{*PWi*}** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Submission**

Submit an electronic copy of your homework solution to the [Blackboard](https://learn.kent.edu/).

18