Exercise sheet for Session 3

Uncertain data management

Antoine Amarilli

December 7th, 2015

1 Exercise 1: From BID to pc-instances

The Bureau for Invasion and Domination in Westeros, using cutting-edge machine learning techniques, has determined an uncertain forecast of the future political state of key cities in the Seven Kingdoms, represented as the following BID instance. (The absence of a city in the table means that the city was sacked and burnt to the ground.)

Westeros

city	**house**	
Winterfell	Stark	0.3
Winterfell	Greyjoy	0.2
Winterfell	Bolton	0.4
King's Landing	Lannister	0.5
King's Landing	Baratheon	0.2
King's Landing	Targaryen	0.2
King's Landing	Stark	0.1

Question 1. Suppose that we know that King's Landing fell to house Stark. What is the probability, knowing this information, that Winterfell is also controlled by house Stark?

Question 2. Using the technique shown in class, write a pc-table that represents the same probabilistic relation.

Question 3. Describe in plain English the meaning of the variables occurring in the resulting pc-instance for the tuples containing "Winterfell".

Question 4. Write in the relational algebra, then in the relational calculus, a query Q_1 that tests whether there is some house that controls two different cities.

Question 5. What is the probability that the pc-instance satisfies Q_1? How many possible worlds of the pc-instance instance satisfy Q_1? How many valuations of the pc-table satisfy Q_1?

Question 6. Write in the relational algebra, then in the relational calculus, a query Q_2 that computes which houses control some city.

Question 7. Construct a pc-table to represent the output of the query Q_2(Westeros). How many possible worlds does it have? Compute the probability that house Stark holds some city.

Question 8. Prove that Q_2(Westeros) cannot be represented by a TID instance.

Question 9. Prove that Q_2(Westeros) cannot be represented by a BID instance.

2 Exercise 2: Adding probabilities to a c-table

Consider the (slightly modified) Boolean c-table obtained in the first question of Exercise 3 for Session 2:

Classes

session	**date**	**prof**	**room**	
2	Nov 30	Antoine	C017	
3	Dec 7	Antoine	C47	$\neg x_1$
4	Dec 14	Silviu	C47	$\neg x_1 \land \neg x_3'$
5	Jan 4	Silviu	C47	$\neg x_1 \land x_2$
6	Jan 11	Silviu	C47	$\neg x_1 \land x_2$

Recall also the semantics of the events:

- x_1: Room C47 collapses. All UDM classes in room C47 must be canceled.
- x_2: D&K students accept to return from vacation. If this does *not* happen, all UDM classes in January are cancelled.
- x_3': Silviu is sick on December 14, we must cancel this class.

Question 1. Assign the following probabilities that the variables are true:

- x_1: 0.01
- x_2: 0.2
- x_3: 0.1

Consider the resulting pc-instance. For each tuple, compute the probability that this tuple is present, i.e,. the total mass of the possible worlds where it occurs. Construct the TID instance Classes2 on the same tuples, where each tuple carries this probability.

Question 2. Consider the Boolean query Q that asks whether Silviu teaches a class. Write it in the relational algebra and in the relational calculus.

Question 3. Consider the result of evaluating Q on the pc-table. What would be the Boolean formula that would annotate the one empty tuple of this result?

Question 4. What is the probability of this Boolean formula? What to conclude about Q on the pc-instance?

Question 5. How does this compare to the probability that the TID instance satisfies query Q? What to conclude about the pc-instance and the TID instance?

Question 6. We observe that the December 7th class has taken place. We formally define the distribution on possible worlds *conditioned* by this observation, where the universe of possible worlds consists of those consistent with the observation, and the probability of each world is its probability in the initial instance, divided by the total probability of the possible world of the initial instance that satisfy the observation.

Write a pc-instance describing the conditioned distribution.