SPLEX TME 9

Bayesian Networks for Clinical Data Analysis

The goal of the TME is to get skills in using the PyAgrum library
`http://agrum.gitlab.io/pages/pyagrum.html`
and to learn how to construct Bayesian networks and dynamic Bayesian networks.

Data (both data sets will be provided during the TME)

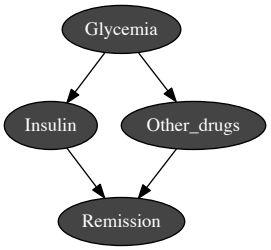
- Prediction of Type 2 Diabetes Remission: dynamic data
- Host and Environmental Data of Obese Patients

You will need to load at least the following packages:

```
import pandas
import numpy
import pyAgrum as gum
import pyAgrum.lib.notebook as gnb
import pyAgrum.lib.bn2graph as bnPlot
import pyAgrum.lib.dynamicBN as gdyn
```

Analysis

1. **Creating a network using PyAgrum.** We would like to model the problem of type 2 diabetes remission after a gastric by-pass surgery which can be represented by the following graph (note, that the problem is simplified extremely!):



The level of glycemia has an important impact on whether insulin or other glucose lowering agents are taken or not. The diabetes remission heavily depends on the number of drugs taken.

The conditional probability tables (CPT) associated with each node of the network are given as follows:

Glycemia > 6.5?

Yes	No
0.5	0.5

	Insulin taken?	
	Yes	No
Glycemia > 6.5		
Yes	0.9	0.1
No	0.1	0.9

	Other drugs taken?	
	Yes	No
Glycemia > 6.5		
Yes	0.75	0.25
No	0.3	0.7

		Remission	
		Yes	No
Insulin	Other drugs		
Yes	Yes	0.1	0.9
No	Yes	0.6	0.4
Yes	No	0.3	0.7
No	No	0.9	0.1

- Create an empty BN network and add nodes
- Create the arcs and the probability tables
- Visualize the graph
- Perform inference (with `LazyPropagation()`)
- Perform inference with evidence. What is the probability to get the remission if the glycemia level is less than 6.5 and no drugs are taken?
- What is the probability to get the remission if the glycemia level is bigger than 6.5 and insulin is prescribed?

The following example can help you:

`http://www-desir.lip6.fr/~phw/aGrUM/docs/last/notebooks/01-tutorial.ipynb.html`

2. **Construct Bayesian networks from real data**. The data are in `SPLEX_env.txt` and `SPLEX_host.txt` files. Construct one network for the environmental variables, one for the host variables, and one with both environmental and host data.

   - Load and discretize the data (the Bayesian networks are learned from discrete data only)
     To discretize the data, each column into 5 bins, you can:

     ```
     l=[]
     for col in data.columns.values:
             bins = numpy.linspace(min(data[col]), max(data[col]), 5)
             l.append(panda.DataFrame(numpy.digitize(data[col], bins),columns=[col]))

     discr_data=panda.concat(l, join='outer', axis=1)
     discr_data.to_csv("data/discr_data.csv",index=False)
     ```

   - Run a learner to learn a networks (test `useLocalSearchWithTabuList()` and `useGreedyHillClimbing()` functions)

     ```
     learner=gum.BNLearner("data/discr_data.csv")
     learner.useLocalSearchWithTabuList()
     bn=learner.learnBN()
     ```

   - Save the obtained networks (you can use `dotize(bn, filename, format='pdf')` function)

- Are the networks learned with `useLocalSearchWithTabuList()` and `useGreedyHillClimbing()` similar?

3. **Dynamic Bayesian networks.** Load data from `dynamic.txt`. In this file, you have HbA1C (glycated hemoglobin), Gly (glycemia), Poids (weight of patients), and Status (remission, non-remission, or partial remission) for time 0, 1 and 5 years after the surgery. Construct a dynamic network to explore temporal dependencies in the data.

   - The first step is to learn a Bayesian network `bn_dynamic` as in the previous task
   - Visualize the network with time slices

     ```
     gdyn.showTimeSlices(bn_dynamic,format="png")
     ```

   An example of how to construct dynamic Bayesian networks in PyAgrum:

   `http://www-desir.lip6.fr/~phw/aGrUM/docs/last/notebooks/32-dynamicBn.ipynb.html`

You will find more examples here
`http://www-desir.lip6.fr/~phw/aGrUM/docs/last/notebooks/`