Computational analysis of biological structures and networks

Exam details

Manuele Bicego

Computer Science Dept.
University of Verona

Workshop

Suggestions

- Suggested structure of the talk:
  - Introduction to the problem (motivation of the study)
  - Main idea of the proposed study (try to avoid formulas unless fundamental)
  - Experimetal results and discussion
  - Conclusion
- Time: 10 minutes (strict!)
- An index of the presentation may help the listeners

Workshop: evaluation criteria

- Relevance to the workshop topic
- Clarity of presentation
- Level of understanding of the paper
- Level of details
- Time (10 minutes)
- Slides (readability and absence of mistakes)
- English
- Answer to possible questions

Written exam: example

- **Question 1** [4 points].

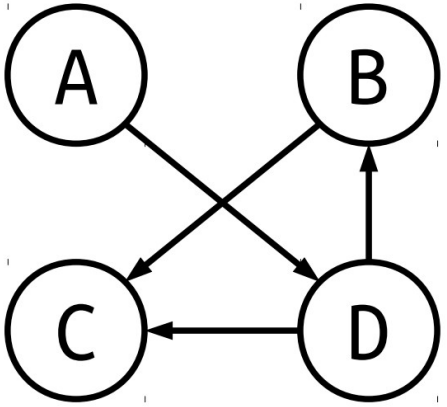
  Describe the main ingredients and the main property of a Bayesian Network, detailing the difference between learning and inference

- **Question 2** [4 points].

  Describe the main ideas behind the Dissimilarity-based representation paradigm, detailing advantages and disadvantages

- **Question 3** [4 points].

  Provide the factorization of the joint probability P(A,B,C,D) defined by the following Bayesian Network. Sketch also how P(B,C|A,D) can be computed from the given Bayesian Network: can we get a formulation without A?



P(A,B,C,D) = ?
P(B,C|A,D) = ?

- **Question 4** [3 points].

  Does this matlab code implement a linear approach to dimensionality reduction applied to the input object X? Please provide a motivation. Which is the final dimension?

```
X = [1; 0.5; 7; 5];  % input vector
Y(1) = 2*X(1) + 2*X(2) + X(3)*X(4);
Y(2) = 3*X(3) + sqrt(2)*X(4);
```

- Answer:
  - NO, this is not a linear approach to dimensionality reduction
  - Motivation: in a linear transformation the new dimensions are obtained as a linear combination of the input dimensions. In `Y(1)` there is a non linear term (`X(3)*X(4)`).
  - The final dimension is 2.

- **Summary**:
  - Three questions on theory and/or exercises
  - One question on code understanding
- **Suggestion 1**:
  - Try to be precise and complete!

Example:
Question: Describe the main ingredients of a Bayesian Network
Your answer:

A set of nodes in the BN graph
A set of edges between the nodes

Not good: Not complete/precise:
- the third ingredient (the conditional probabilities) is missing
- important information unspecified: each node represents a random variable and the edges are such that the resulting graph is acyclic

- **Suggestion 2**: try to avoid to write concepts which are not related to the question

Example:
Question: Describe the main ingredients of a Bayesian Network
In your answer: […….] + "An important rule, sometimes used when using BN, is the following: P(A,B) = P(A|B) P(B)"

This last concept is not necessary to define the ingredients of the BN

- Note: if the unrelated information is "drastically" uncorrect, you can get a penalization

Example:
In your answer: […….] + "An important rule, sometimes used when using BN, is the following: P(A,B) = P(A) + P(B)"

The formula is drastically uncorrect, you can get a penalization