732A96/TDDE15 ADVANCED MACHINE LEARNING

EXAM 03/01-2018

TEACHERS

Jose M. Peña. Phone: 0700895280.
Mattias Villani. Will visit the room.

GRADES

- For 732A96 (A-E means pass):
  - A=19-20 points
  - B=17-18 points
  - C=12-16 points
  - D=10-11 points
  - E=8-9 points
  - F=0-7 points
- For TDDE15 (3-5 means pass):
  - 5=18-20 points
  - 4=12-17 points
  - 3=8-11 points
  - U=0-7 points

The total number of points is rounded to the nearest integer. In each question, full points requires clear and well motivated answers.

ALLOWED MATERIAL

Hard copy of Bishop's book, and the content of the folder given_files in the exam system.

INSTRUCTIONS

The answers to the exam should be submitted in a single PDF file using the communication client. You can make a PDF from LibreOffice (similar to Microsoft Word). You can also use Markdown from RStudio. Include important code needed to grade the exam (inline or at the end of the PDF file). Submission starts by clicking the button "Skicka in uppgift" in the communication client. Then, follow the instructions. Note that the system will let you know that the exam has been submitted, but will not tell you that it was received. This is ok and your solution has actually been received.

Do not ask question through the communication client. The teachers will be reachable by phone, and they will visit the room too.

1. GRAPHICAL MODELS (5 P)

- Learn a Bayesian network (BN) from 80 % of the Asia dataset. The dataset is included in the `bnlearn` package. To load the data, run `data("asia")`. Learn both the structure and the parameters. Use any learning algorithm and settings that you consider appropriate. Use the BN learned to classify the remaining 20 % of the Asia dataset in two classes: $S = yes$ and $S = no$. In other words, compute the posterior probability distribution of S for each case and classify it in the most likely class. To do so, you may want to use exact or approximate inference with the help of the `bnlearn` and `gRain` packages. Report the confusion matrix, i.e. true/false positives/negatives. (3 p)
- In the previous exercise, you classified the variable S given observations for all the rest of the variables. Now, you are asked to classify S given observations only for the so-called Markov blanket of S, i.e. its parents plus its children plus the parents of its children minus S itself. Report again the confusion matrix. (1 p)
- Explain why the results of the previous exercises coincide (as long as you use the same BN learned). (1 p)

2. HIDDEN MARKOV MODELS (5 P)

- Use the `HMM` package to implement the dishonest casino hidden Markov model (HMM). Recall that this HMM is included in the package and, thus, you can get a description of it from the package. Use any transition and emission probabilities that you consider appropriate. Then, modify your implementation so that when a die is chosen, it is used for at least three consecutive throws. In particular, this regime's minimum duration should be implemented implicitly by duplicating hidden states and the emission model, i.e. do not use increasing or decreasing counting variables. Finally, sample the HMM built. (3 p)
- Explain how to learn the parameters of a HMM given a sample of observations, i.e. no hidden states are included in the sample. Please, be as detailed as possible but do not use more than 250 words. (2 p)

3. GAUSSIAN PROCESSES (5 P)

See attached file `Exam732A96_180103Question3.pdf`.

4. State Space Models (5 p)

- Propose your own state space model (SSM). It has to be significantly different from the SSMs used in the labs and in previous exams, i.e. do not simply change the numbers in some existing equations, change the equations as well. Implement the particle filter for the proposed SSM. Show that it works correctly, e.g. proceed as you did in the corresponding lab. (4 p)
- Name the main differences between the particle filter and the Kalman filter. (1 p)

Good luck !