732A96/TDDE15 ADVANCED MACHINE LEARNING

EXAM 2020-08-26

Teacher

Jose M. Peña. Available by e-mail.

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.

Instructions

This is an individual exam. No help from others is allowed. No communication with others regarding the exam is allowed. You are allowed to use the course literature, the course slides and your own notes. Answers to the exam questions may be sent to Urkund.

The answers to the exam should be submitted in a single PDF file using LISAM. 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). The exam is anonymous, i.e. do not write your name anywhere.

1. Graphical Models (8 p)

(1) (2 p) Consider the following directed and acyclic graph:

$$\begin{tikzcd} & C \arrow[d] \arrow[ddl] \arrow[ddr] & \\ & D & \\ A \arrow[rr] & & Y \end{tikzcd}$$

Use `bnlearn` to implement it. Assume that all the random variables are binary.

(2) (1 p) Explain your solution to the previous exercise. Refer to the code produced.

(3) (4 p) You are asked to randomly parameterize 1000 times the graph constructed in the previous exercise. Each random parameterization should be obtained by drawing the parameters values for each conditional probability table from a uniform distribution. For each random parameterization, you should check whether $p(y|a,c)$ and $p(y|a,d)$ are monotone in C and D, respectively. See the definitions below. You are asked to report how many of the 1000 random parameterizations result in (i) $p(y|a,c)$ is monotone in C but $p(y|a,d)$ is not monotone in D, and (ii) $p(y|a,d)$ is monotone in D but $p(y|a,c)$ is not monotone in C. To solve this exercise, you must use `gRain`. You may want to use the function `as.grain` to convert the `bnlearn` object from the previous exercise into a `gRain` object.

We say that $p(y|a,c)$ is nondecreasing in C if

$$p(Y=1|A=1,C=1) \geq p(Y=1|A=1,C=0) \text{ and } p(Y=1|A=0,C=1) \geq p(Y=1|A=0,C=0).$$

Likewise, $p(y|a,c)$ is nonincreasing in C if

$$p(Y=1|A=1,C=1) \leq p(Y=1|A=1,C=0) \text{ and } p(Y=1|A=0,C=1) \leq p(Y=1|A=0,C=0).$$

Moreover, $p(y|a,c)$ is monotone in C if it is nondecreasing or nonincreasing in C. We define similarly that $p(y|a,d)$ is monotone in D.

(4) (1 p) Explain your solution to the previous exercise. Refer to the code produced.

2. Hidden Markov Models (5 p)

You are asked to build a hidden Markov model (HMM) to model a weather forecast system. The system is based on the following information. If it was rainy (respectively sunny) the last two days, then it will be rainy (respectively sunny) today with probability 0.75 and sunny (respectively rainy) with probability 0.25. If the last two days were rainy one and sunny the other, then it will be rainy today with probability 0.5 and sunny with probability 0.5. Moreover, the weather stations that report the weather back to the system malfunction with probability 0.1, i.e. they report rainy weather when it is actually sunny and vice versa. Implement the weather forecast system described using the `HMM` package. Sample 10 observations from the HMM built.

Hint: You may want to have hidden random variables with four states encoding the weather not only today but also yesterday. However, your observed random variables should have two states corresponding to the weather today.

3. Gaussian Processes (7 p)

(1) (2 p) The file `KernelCode.R` distributed with the exam contains code to construct a `kernlab` function for the Matern covariance function with $\nu = 3/2$:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{\ell}\right) \exp\left(-\frac{\sqrt{3}r}{\ell}\right)$$

where $r = |\mathbf{x} - \mathbf{x}'|$. Let $f \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$ a priori and let $\sigma_f^2 = 1$ and $\ell = 0.5$. Plot $k(0, z)$ as a function of z. You can use the grid `zGrid = seq(0.01,1,by=0.01)` for the plotting. Interpret the plot. Connect your discussion to the smoothness of f. Finally, repeat this exercise with $\sigma_f^2 = 0.5$ and discuss the effect this change has on the distribution of f.

(2) (2 p) You are asked to extend your work on Lab 4 with the Tullinge temperatures. In the lab, you predicted the temperature as function of time with the following hyperparameter values: $\sigma_f = 20$ and $\ell = 0.2$. Now, you are asked to search for the best hyperparameter values by maximizing the log marginal likelihood. You may want to check the corresponding slides for the theoretical details. Recall that you implemented Algorithm 2.1 in the book by Rasmussen and Williams, which already returns the log marginal likelihood. Use a grid search to search for the best hyperparameter values, i.e. try different hyperparameter values while time permits.

(3) (2 p) You are asked to extend your work on Lab 4 with the banknote fraud data. In the lab, you used the default squared exponential kernel (a.k.a. radial basis function) with automatic hyperparameter value determination. Now, you are asked to search for the best hyperparameter value by using a validation dataset. Use the four covariates to classify. Use a grid search, i.e. try different hyperparameter values while time permits.

(4) (1 p) Explain your solution to the previous exercise. Refer to the code produced.