732A96/TDDE15 ADVANCED MACHINE LEARNING

EXAM 2018-10-24

Teacher

Jose M. Peña. Will visit the room. Available by phone too.

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 teacher will be reachable by phone, and he will visit the room too.

1. Graphical Models (5 p)

Select 4000 data points from the Asia dataset (included in the `bnlearn` package) for training and 1000 points for testing by running the following lines:

```
set.seed(567)
data("asia")
ind <- sample(1:5000, 4000)
tr <- asia[ind,]
te <- asia[-ind,]
```

Use the first 10, 20, 50, 100, 1000, 2000 of the training points to learn a naive Bayes (NB) classifier for the random variable S. Recall that a NB classifier only has directed edges from S to the rest of the nodes. Report the accuracy of the classifier on the 1000 test points. You have to create the NB classifier by hand, i.e. you are not allowed to use the function `naive.bayes` from the `bnlearn` package. When learning the parameters for the classifier, use `method="bayes"`. Ignore the warnings. When classifying, use the `gRain` package for exact inference. Classify according to the most likely class label.

Repeat the exercise above for the classifier resulting from reversing the edges in the NB classifier, i.e. the edges go now from the rest of the nodes to S. Compare the results you obtain for both classifiers, and explain why they may differ.

2. Hidden Markov Models (5 p)

Implement the forward phase of the forward-backward algorithm as it appears in the course slides or in the book by Bishop. Run it on the data that you used in the lab on hidden Markov models. Compute the accuracy of the filtered distributions. Show that you obtain the same accuracy when using the `HMM` package.

3. State Space Models (5 p)

Consider the following state space model (SSM):

$$
\begin{aligned}
p(x_t|x_{t-1}) &= \mathcal{N}(x_t|x_{t-1}+1, 1) \\
p(z_t|x_t) &= \mathcal{N}(z_t|x_t, 5) \\
p(x_0) &= \mathcal{N}(x_0|50, 10)
\end{aligned}
$$

Implement and simulate the SSM above for $T = 100$ time steps to obtain a sequence of observations $z_{1:T}$ and hidden states $x_{1:T}$. Implement the Kalman filter as it appears in the course slides or in the book by Thrun et al. Note that the SSM above specifies standard deviations 1, 5 and 10 for the transition, emission and initial models. However, the Kalman filter in the slides and in the book is described in terms of variances instead. Since there are no actions in the SSM above, assume $u_t = 1$.

Run the Kalman filter on the observations $z_{1:T}$. Report the mean and standard deviation of the errors for the $T = 100$ time steps. The error for time t is defined as $abs(x_t - E[x_t])$, where the expectation is with respect to the filtered distribution (a.k.a. belief function).

4. Gaussian Processes (5 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 $r = |\mathbf{x} - \mathbf{x}'|$.

(1) 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 p)

(2) The file `lidar.RData` distributed with the exam contains two variables *logratio* and *distance*. Load the variables into memory with the R command `load("lidar.RData")`. Compute the posterior distribution of f in the model

$$logratio = f(distance) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 0.05^2).$$

You should do this for both length scales $\ell = 1$ and $\ell = 5$. Set $\sigma_f = 1$. Your answer should be in the form of a scatter plot of the data overlayed with curves for (a) the posterior mean of f, (b) 95 % probability intervals for f, and (c) 95 % prediction intervals for y. Use the `gausspr` function in the `kernlab` package for (a), but not for (b) and (c) since the function seems to contain a bug. For (b) and (c) instead, find the appropriate expression in the course slides or in the book by Rasmussen and Williams and implement it. You are not allowed to use Algorithm 2.1. Discuss the differences in results from using the two length scales. (3 p)