732A96 ADVANCED MACHINE LEARNING

EXAM

JOSE M. PEÑA
IDA, LINKÖPING UNIVERSITY, SWEDEN

1. Graphical Models (5 p)

- Learn a Bayesian network (both structure and parameters) from the Asia dataset that is distributed in the `bnlearn` package. Use any learning algorithm and settings that you consider appropriate. Identify one d-separation in the Bayesian network learned and show that it indeed corresponds to a probabilistic independence statement. To do so, you may want to use the functions in the `gRain` package. (2 p)
- There are 29281 directed and acyclic graphs with five nodes. Compute approximately the fraction of these 29281 graphs that represent an independence model that can be represented with a Markov network. You may want to use the function `skeleton` of the `bnlearn` package, which outputs the undirected graph that results from dropping the directions of the edges in the input graph. (2 p)
- Explain how to perform probabilistic reasoning (i.e. compute a conditional probability distribution) in a Bayesian network. Please, be as detailed as possible but do not use more than 250 words. (1 p)

2. Hidden Markov Models (5 p)

- Recall Lab 2 where you were asked to build a HMM for modeling the behavior of a robot that walked around a ring. The ring was divided into 10 sectors. At any given time point, the robot was in one of the sectors and decided with equal probability to stay in that sector or move to the next sector. You did not have direct observation of the robot. However, the robot was equipped with a tracking device that you could access. The device was not very accurate though: If the robot was in the sector i, then the device reported that the robot was in the sectors [i - 2, i + 2] with equal probability.

  You are now asked to extend the HMM built in Lab 2 as follows. The observed random variable has now 11 states, corresponding to the 10 sectors of the ring plus a 11th state to represent that the tracking device is malfunctioning. If the robot is in the sector i, then the device will report that it is malfunctioning with probability 0.5 and that the robot is in the sectors [i - 2, i + 2] with probability 0.1 each. Implement the extension just described by using the `HMM` package. Moreover, consider the observations 1, 11, 11, 11, i.e. the tracking device reports sector 1 first, and then malfunctioning for three time steps. Compute the most probable path using the smoothed distribution and the Viterbi algorithm. Explain why the paths differ, if they do. (2 p)
- You are asked to build a 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 in two consecutive days. (2 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. (1 p)

LINKÖPING UNIVERSITY
Dept. of Computer and Information Science
Division of Statistics and Machine Learning
Jose M. Peña and Mattias Villani

2016-10-20
Advanced Machine Learning, 6 hp
732A96

Computer Exam - Advanced Machine Learning (732A96), 6 hp

Time:	2-6 PM
Allowable material:	- Paper copy of the book 'Pattern Recognition and Machine Learning' by Bishop. - 100 page PDF file set up by the student (readable at full scale) available through the computer exam system. - Slides from the lectures, available through the computer exam system.
Teachers:	Jose M. Peña. Phone: $070 - 0895280$ (available from 2 PM) Mattias Villani. Phone: $070 - 0895205$ (available from 3.15 PM)
Grades:	Maximum number of credits on the exam: 20. Maximum number of credits on each exam question: 5. A=19-20 points B=17-18 points C=12-16 points D=10-11 points E=8-9 points F=0-7 points The total exam score is rounded to the nearest integer.

Full credit requires clear and well motivated answers.

1. Graphical Models

   (a) Learn a Bayesian network (both structure and parameters) from the Asia dataset that is distributed in the bnlearn package. Use any learning algorithm from the bnlearn package and settings that you consider appropriate. Use the Bayesian network learned to compute the conditional probability of person having visited Asia given that the person has bronchitis and the X-rays came positive, i.e. $p(A|X = TRUE, B = TRUE)$. Use both the approximate and exact methods. (2 p)

   (b) There are 29281 DAGs with five nodes. Compute approximately the fraction of the 29281 DAGs that represent an independence model that can be represented with a Markov network. You may want to use the function `skeleton` of the bnlearn package, which outputs the undirected graph that results from dropping the directions of the edges in the input graph. (2 p)

   (c) Explain how to perform probabilistic reasoning (i.e. compute a conditional probability distribution) in a Bayesian network. Please, be as detailed as possible but do not use more than 250 words. (1 p)

2. HIDDEN MARKOV MODELS

Recall Lab 2 where you were asked to build a HMM for modeling the behavior of a robot that walks around a ring. The ring is divided into 10 sectors. At any given time point, the robot is in one of the sectors and decides with equal probability to stay in that sector or move to the next sector. You do not have direct observation of the robot. However, the robot is equipped with a tracking device that you can access. The device is not very accurate though: If the robot is in the sector i, then the device will report that the robot is in the sectors $[i-2, i+2]$ with equal probability.

(a) You are asked to extend the HMM built in Lab 2 as follows. The observed random variable has now 11 states, corresponding to the 10 sectors of the ring plus a 11th state to represent that the tracking device is malfunctioning. If the robot is in the sector i, then the device will report that it is malfunctioning with probability 0.5 and that the robot is in the sectors $[i-2, i+2]$ with probability 0.1 each. Implement the extension just described by using the HMM package. Moreover, consider the observations $1, 11, 11, 11$, i.e. the tracking device reports sector 1 first, and then malfunctioning for three time steps. Compute the most probable path using the smoothed distribution and the Viterbi algorithm. Explain why the paths differ, if they do. (2 p)

(b) You are asked to modify the HMM built in Lab 2 as follows. The ring has now only five sectors. If the robot is in the sector i, then the tracking device will report that the robot is in the sectors $[i-1, i+1]$ with equal probability. The rest of the sectors receive zero probability. The robot now spends at least two time steps in each sector. You are asked to implement this modification. In particular, the regime's minimum duration should be implemented implicitly by duplicating hidden states and the observation model, i.e. do not use increasing or decreasing counting variables. (2 p)

(c) 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. (1 p)

3. GAUSSIAN PROCESSES

The file KernelCode.txt distributed with the exam contains code to construct the following three kernel functions in the kernlab format:

$$k_1(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{r^2}{2\ell^2}\right)$$

$$k_2(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \left(1 + \frac{r^2}{2\alpha\ell^2}\right)^{-\alpha}$$

$$k_3(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{\ell}\right) \exp\left(-\frac{\sqrt{3}r}{\ell}\right).$$

where $r = |\mathbf{x} - \mathbf{x}'|$.

(a) Now, let $\sigma_f = \ell = 1$ and assume a prior mean of zero for the process. Compute and plot the covariance function corresponding to the three different kernel functions k_1, k_2, k_3. Use three different values of α: 1/2, 2 and 20 in k_2. Plot all covariance functions in the same graph over the domain `r<-seq(0,4,by=0.01)`. Argue about the dependence properties of the Gaussian process from these different covariance functions. For example, what do you expect simulated realizations from the different kernel function to look like? What does this graph tell you about the effect of the hyperparameter α on the process? (1.5 p)
[Hint 1: the function `kernelMatrix` in kernlab may be useful, but is not strictly necessary to solve this exercise. Hint 2: you may find it useful to simulate realizations of the process for the different kernels. This is not required to get full score, but may help you when arguing about the properties of the covariance functions].

(b) The file GPdata.RData contains two variables `y` and `x`. Load the variables into memory with the R command `load("GPdata.RData")`. Compute the posterior distribution of f in the model

$$y = f(x) + \varepsilon, \quad \varepsilon \sim N(0, 0.5^2).$$

Use both the squared exponential (k_1) kernel and the k_3 kernel with $\ell = 1$ and $\sigma_f = 1$. Your answer should be in the form of a scatter plot of the data overlayed with curves for

i. the posterior mean of f

ii. 95% probability intervals for f

iii. 95% prediction intervals for a new data point y

Interpret the results under i)-iii), and explain the difference between the results from ii) and iii). Discuss the differences in results from using the two kernels. Would you prefer one kernel over the other?

Use the `gausspr` function in the kernlab package for i), but not for ii) and iii).

[Hint: $Cov(f) = K(X_\star, X_\star) - K(X_\star, X)\left[K(X, X) + \sigma_n^2 I\right]^{-1} K(X, X_\star)$ and remember that `%*%` does matrix multiplication and `solve` computes inverses in R] (2 p)

(c) (No need to do any computations here). Discuss how a Bayesian would handle the case where the kernel hyperparameters are unknown. Discuss how Bayesian analysis can be used to choose between the kernels k_1, k_2 and k_3. (1.5 p)

4. State-space models

The data set radiation_data.Rda contains $T = 365$ daily measurements of radioactivity (unit: micro sieverts per hour) between December 5, 2012 and December 4, 2013 from the Township village center near the Fukushima power plant. Assume the following local linear trend model for the data

$$\begin{aligned}
y_t &= \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2 = 0.16) \\
\alpha_t &= \alpha_{t-1} + \beta_{t-1} + v_t^{(1)}, \quad v_t^{(1)} \sim N\left(0, \sigma_{v^{(1)}}^2 = 0.035\right) \\
\beta_t &= \beta_{t-1} + v_t^{(2)}, \quad v_t^{(2)} \sim N\left(0, \sigma_{v^{(2)}}^2 = 3.06 \cdot 10^{-12}\right),
\end{aligned}$$

where ϵ_t, $v_t^{(1)}$ and $v_t^{(2)}$ are independent. Assume a priori that α_0 and β_0 are independent with priors $p(\alpha_0) = N(10, 10^2)$ and $p(\beta_0) = N(0, 10^2)$.

You should use the R package `dlm` to solve this problem. Load the data frame by the command `load("Radiation_data.Rda")`. The variable of interest (y_t) is `dose` which you can access by `Radiation_data$dose`. Only make plots of the results when asked to (so you don't waste your valuable time!).

(a) Let the state vector be $x_t = (\alpha_t, \beta_t)^T$ and formulate the model as a dynamic linear Gaussian state-space model. That is, write

$$\begin{aligned}
y_t &= F x_t + \epsilon_t, \quad \epsilon_t \sim N(0, V) \\
x_t &= G x_{t-1} + v_t, \quad v_t \sim N(0, W) \\
p(x_0) &= N(\mu_0, \Sigma_0)
\end{aligned}$$

and find the system matrices F, G, the variance components V, W, and the prior mean μ_0 and covariance Σ_0 of the state vector at $t = 0$. (1 p)

(b) Compute the filtering distribution $p(x_t|y_{1:t})$ and the smoothing distribution $p(x_t|y_{1:T})$, for $t = 1, ..., T = 365$. For the local level parameter (α_t), plot the expected value of its filtering distribution, i.e. $\mathbb{E}\left(\alpha_t|y_{1:t}\right)$. (1 p)

(c) Find the distribution of the maximum of the local slope component (β_t) over the observed period given all data. Use e.g. a histogram or kernel density estimator to present your result. [Hint: The distribution of interest is

$$p\left(\max\left\{\beta_{1:T}\right\}|y_{1:T}\right),$$

which is analytically intractable but can be computed by simulation.] (2 p)

(d) Forecast (use `dlmForecast`) the daily radiation level for the period December 5, 2013 - December 24, 2013. What is the probability that the radiation level is lower than 7 micro sieverts on Christmas Eve 2013 (December 24)? (1 p)

Good luck!

Jose and Mattias

732A96/TDDE15 ADVANCED MACHINE LEARNING

EXAM 19/10-2017

TEACHERS

Jose M. Peña. Phone: 0700895280.

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 the Asia dataset that 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. Identify a d-separation in the BN learned and show that it indeed corresponds to an independence in the probability distribution represented by the BN. To do so, you may want to use exact or approximate inference with the help of the `bnlearn` and `gRain` packages. (2.5 p)
- There are 29281 directed and acyclic graphs (DAGs) with five nodes. Compute approximately the fraction of these 29281 DAGs that are essential. An essential DAG is a DAG that is not Markov equivalent to any other DAG. The simplest way to solve the exercise may be by re-using the code that you produced for the lab. For this to work, you have to figure out how to determine if a DAG is essential by just looking at its CPDAG (a.k.a. essential graph). (2.5 p)

2. Hidden Markov Models (5 p)

- Consider a robot moving along a straight corridor. The corridor is divided into 100 segments. The corridor has three doors: The first spans segments 10, 11 and 12, the second spans segments 20, 21 and 22, and the third spans segments 30, 31, and 32. In each time step, the robot moves to the next segment with probability 0.9 and stays in the current segment with probability 0.1. You do not have direct observation of the robot. However, the robot is equipped with a sensor that is able to detect whether the robot is or is not in front of a door. The accuracy of the sensor is 90 %. Initially, the robot is in any of the 100 segments with equal probability. You are asked to build a hidden Markov model (HMM) to model the robot's behavior. You may want to use the `HMM` package. (2.5 p)
- Give a sequence of observations $x_{1:t}$ such that $p(z_t|x_{1:t})$ is unimodal, i.e. there is a segment that is more likely to contain the robot than the rest of the segments. Recall that initially the robot is in any of the 100 segments with equal probability. Use the `HMM` package to show that the sequence really solves the exercise. (2.5 p)

  Note that the function `which.max` only returns the first maximum, not all of them. Thus, you may want to use the function below, which does return all the maxima.

```
which.maxima<-function(x){
  return(which(x==max(x)))
}
```

3. Gaussian Processes

The file `KernelCode.R` distributed with the exam contains code to construct a `kernlab` function for the Squared Exponential covariance function:

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

where $r = |\mathbf{x} - \mathbf{x}'|$.

(a) Let $f \sim GP(0, k(\mathbf{x}, \mathbf{x}'))$ a priori and let $\sigma_f^2 = 1$. Simulate and plot 5 realizations from the prior distribution of f over the grid `xGrid = seq(-1,1,by=0.1)` for the two length scales $\ell = 0.2$ and $\ell = 1$ (in two separate figures). Compute

i. $\mathrm{Corr}(f(0), f(0.1))$ and
ii. $\mathrm{Corr}(f(0), f(0.5))$
for the two length scales. Discuss the results and connect your discussion to the concept of smoothness of f. [Hint: you will probably need the `mvtnorm` package.] (2 p)

(b) The file GPdata.RData contains two variables `y` and `x`. Load the variables into memory with the R command `load("GPdata.RData")`. Compute the posterior distribution of f in the model

$$y = f(x) + \varepsilon, \quad \varepsilon \sim N(0, 0.2^2).$$

You should do this for both length scales $\ell = 0.2$ and $\ell = 1$. Set $\sigma_f = 1$. Your answer should be in the form of a scatter plot of the data overlayed with curves for

i. the posterior mean of f
ii. 95% probability intervals for f
iii. 95% prediction intervals for a new data point y
Explain the difference between the results from ii) and iii). Discuss the differences in results from using the two length scales. Do you think a GP with a squared exponential kernel is a good model for this data? If not, why?
Use the `gausspr` function in the kernlab package for i), but not for ii) and iii).
[Hint: $Cov(f) = K(X_\star, X_\star) - K(X_\star, X)\left[K(X, X) + \sigma_n^2 I\right]^{-1} K(X, X_\star)$ and remember that `%*%` does matrix multiplication and `solve` computes inverses in R] (2 p)

(c) (No need to do any computations here). Discuss how a Bayesian would handle the case where the kernel hyperparameters are unknown. What if the noise variance is unknown? (1 p)

4. State Space Models (5 p)

- Consider the following state space model (SSM):

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

  Implement and simulate the SSM above for $T = 10000$ 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.

  Run the Kalman filter on the observations $z_{1:T}$. Report the mean and standard deviation of the errors for the $T = 10000$ 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). (2 p)
- Repeat the exercise above with the particle filter. You may want to re-use the code you produced for the lab. (2 p)
- Compare the performance (i.e. mean error and runtime) of the Kalman and particle filters on the same sequence of observations when using 10, 50 and 100 particles. Explain the differences in performance between the filters. (1 p)

  To measure runtime, you may want to use the code below.

```
start_time <- Sys.time()
# Your code
end_time <- Sys.time()
end_time - start_time
```

Good luck !

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 !

2. Gaussian Processes

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 $r = |\mathbf{x} - \mathbf{x}'|$.

(a) Let $f \sim GP(0, k(\mathbf{x}, \mathbf{x}'))$ a priori and let $\sigma_f^2 = 1$. Plot $k(0, 0.1)$ as a function of the hyperparameter ℓ. You can use the grid `ell = seq(0.01,1,by=0.01)` for the plotting. Carefully interpret this plot in a way that shows that you understand what it shows. Connect your discussion to the smoothness of f. (2.5 p.)

(b) The file GPdata.RData contains two variables `y` and `x`. Load the variables into memory with the R command `load("GPdata.RData")`. Compute the posterior distribution of f in the model

$$y = f(x) + \varepsilon, \quad \varepsilon \sim N(0, 0.15^2).$$

You should do this for both length scales $\ell = 0.2$ and $\ell = 1$. Set $\sigma_f = 1$. Your answer should be in the form of a scatter plot of the data overlayed with curves for

i. the posterior mean of f

ii. 95% probability intervals for f

iii. 95% prediction intervals for a new data point y

Use the `gausspr` function in the kernlab package for i), but not for ii) and iii).

Discuss the differences in results from using the two length scales.

Do you think a GP with a Matern($\nu = 3/2$) kernel is a good model for this data? If not, what could be the problem with this model? Can you think of a better model? Discuss.

[Hint: $Cov(f) = K(X_\star, X_\star) - K(X_\star, X)\left[K(X, X) + \sigma_n^2 I\right]^{-1} K(X, X_\star)$ and remember that `%*%` does matrix multiplication and `solve` computes inverses in R] (2.5 p)

732A96/TDDE15 ADVANCED MACHINE LEARNING

EXAM 2018-08-29

TEACHERS

Jose M. Peña. Will visit the room.
Mattias Villani. Available by phone.

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 (both structure and parameters) from the Asia dataset that is distributed in the `bnlearn` package. To load the data, run `data("asia")`. Use any learning algorithm and settings that you consider appropriate. Use the Bayesian network learned to compute the probability distribution of a person having visited Asia given that the person has bronchitis and the X-rays came positive, i.e. $p(A|X = TRUE, B = TRUE)$. Use both the approximate and exact methods. (3 p)
- There are 29281 DAGs with five nodes. Compute approximately the fraction of the 29281 DAGs that represent an independence model that can be represented with a Markov network. You may want to use the function `skeleton` of the `bnlearn` package, which outputs the undirected graph that results from dropping the directions of the edges in the input graph. (2 p)

2. Hidden Markov Models (5 p)

- Recall Lab 2 where you were asked to build a HMM for modeling the behavior of a robot that walked around a ring. The ring was divided into 10 sectors. At any given time point, the robot was in one of the sectors and decided with equal probability to stay in that sector or move to the next sector. You did not have direct observation of the robot. However, the robot was equipped with a tracking device that you could access. The device was not very accurate though: If the robot was in the sector i, then the device reported that the robot was in the sectors [i - 2, i + 2] with equal probability.

  You are now asked to extend the HMM built in Lab 2 as follows. The observed random variable has now 11 states, corresponding to the 10 sectors of the ring plus a 11th state to represent that the tracking device is malfunctioning. If the robot is in the sector i, then the device will report that it is malfunctioning with probability 0.5 and that the robot is in the sectors [i - 2, i + 2] with probability 0.1 each. Implement the extension just described by using the `HMM` package. Moreover, consider the observations 1, 11, 11, 11, i.e. the tracking device reports sector 1 first, and then malfunctioning for three time steps. Compute the most probable path using the smoothed distribution and the Viterbi algorithm. (3 p)
- You are asked to build a 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 in two consecutive days. (2 p)

3. GAUSSIAN PROCESSES (5 P)

See attached file Exam732A96_180829Question3.pdf.

4. State Space Models (5 p)

- Consider the following state space model (SSM):

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

  Implement and simulate the SSM above for $T = 10000$ 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.

  Run the Kalman filter on the observations $z_{1:T}$. Report the mean and standard deviation of the errors for the $T = 10000$ 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). (3 p)
- Repeat the exercise above with the particle filter. You may want to re-use the code you produced for the lab. (2 p)

Good luck !

1.

2. Gaussian Processes

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 $r = |\mathbf{x} - \mathbf{x}'|$.

(a) Let $f \sim GP(0, k(\mathbf{x}, \mathbf{x}'))$ 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. Carefully interpret this plot in a way that shows that you understand what it shows. 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.)

(b) The file lidar.RData 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

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

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

i. the posterior mean of f

ii. 95% probability intervals for f

iii. 95% prediction intervals for a new data point y

Use the `gausspr` function in the kernlab package for i), but not for ii) and iii).
Discuss the differences in results from using the two length scales.
Do you think a GP with a Matern($\nu = 3/2$) kernel is a good model for this data? If not, what could be the problem with this model? Can you think of a better model? Discuss.
[Hint: $Cov(f) = K(X_\star, X_\star) - K(X_\star, X)\left[K(X, X) + \sigma_n^2 I\right]^{-1} K(X, X_\star)$ and remember that `%*%` does matrix multiplication and `solve` computes inverses in R] (2 p)

(c) (No need to do any computations here). Discuss how a Bayesian would handle the case where the kernel hyperparameters are unknown. What if the noise variance is unknown? (1 p)