EINDHOVEN UNIVERSITY OF TECHNOLOGY

Surname and initials:

Student number:

Examination cover sheet

(to be completed by the examiner)

Course name: Bayesian Machine Learning and Information Processing   Course code: 5SSD0

Date: 28-Jan-2021

Start time: 13:30   End time: 16:30 (on campus); 17:15 (online, at home)

Number of pages: part A: 5 pages. Part B: 5 pages

Number of questions: part A: 3 questions. Part B: 3 questions

Maximum number of points/distribution of points over questions: each question 5 points (total 30 points)

Method of determining final grade: grade = (sum of points)/(30) + correction. Correction as discussed in class and at piazza.

Answering style: formulation, order, argumentation, multiple choice: multiple choice

Exam inspection:

Other remarks: good luck!

INSTRUCTIONS FOR STUDENTS AND INVIGILATORS
(to be indicated by examinor)

Write in black or blue. Pencil only allowed for drawings.

Permitted examination aids (to be supplied by students):

- ☐ Computer
- ☐ Calculator Graphic
- ☐ Calculator
- ☐ Lecture notes/book
- ☐ One A4 sheet of annotations
- ☐ Dictionaries. If yes, please specify:

Important:

- an exam >90 minutes consists of a Part A and a Part B. Part A needs to be collected after 90 minutes, before handing out part B. Ensure students are aware of this
- students that started on part B of the exam, are only allowed to hand it in and leave the room after the simultaneously conducted online exam part B has started. This is 11.05 (morning), 15.35 (afternoon)
- examinees are only permitted to visit the toilets under supervision
- it is not permitted to leave the examination room within 15 minutes of the start and within the final 15 minutes of the examination, unless stated otherwise
- examination scripts (fully completed examination paper, stating name, student number, etc.) must always be handed in
- the house rules must be observed during the examination
- the instructions of subject experts and invigilators must be followed
- keep your work place as clean as possible: put pencil case and breadbox away, limit snacks and drinks
- examinees are not permitted to share examination aids or lend them to each other

During written examinations, the following actions will in any case be deemed to constitute fraud or attempted fraud:

- using another person's proof of identity/campus card (student identity card)
- having a mobile telephone or any other type of media-carrying device on your desk or in your clothes
- using, or attempting to use, unauthorized resources and aids, such as the internet, a mobile telephone, smartwatch, smart glasses etc.
- having any paper at hand other than that provided by TU/e, unless stated otherwise
- copying (in any form)
- visiting the toilet (or going outside) without permission or supervision

The final grade will be announced no later than fifteen working days after this examination took place. Final grades of first-year bachelor study components in Q4 will be announced within 5 working days. Final test grades of bachelor study components in the interim period will be announced no later than 5 working days before the 1st of September.

Exercises

1	2	3

Surname, First name

Bayesian machine learning and information processing (5SSD0)

5SSD0 Bayesian machine learning and information processing Q2 20-21 Part A Handwritten exam

1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
0	0	0	0	0	0	0

Question 1

Which of the following statements are true?

1p **1a** It is more appropriate to say "the likelihood of the parameters", than "the likelihood of the data".

(a) true (b) false

1p **1b** If X and Y are independent Gaussian distributed variables, then $Z = 3X - XY$ is also a Gaussian distributed variable.

(a) true (b) false

1p **1c** For a given Linear Gaussian Dynamical System with observations $\{x_t\}$ and latent states $\{z_t\}$, the Kalman filter is a recursive solution to the inference problem $p(z_t|x_{1:t})$, based on a state estimate at the previous time step $p(z_{t-1}|x_{1:t-1})$ and a new observation x_t.

(a) true (b) false

1p **1d** In the context of parameter estimation, maximum likelihood estimation always selects the parameter values where the Bayesian posterior distribution is maximal.

(a) true (b) false

1p **1e** Bayes rule is inconsistent with the Method of Maximum Relative Entropy as a method of inference.

(a) true (b) false

Question 2

Let $\boldsymbol{x}_n \in \mathbb{R}^N$ and $\boldsymbol{z}_n \in \mathbb{R}^M$ with $M << N$. Given is a model

$$
\begin{aligned}
\boldsymbol{x}_n &= \boldsymbol{W} \boldsymbol{z}_n + \boldsymbol{\epsilon}_n \\
\boldsymbol{z}_n &\sim \mathcal{N}(0, \boldsymbol{I}) \\
\boldsymbol{\epsilon}_n &\sim \mathcal{N}(0, \boldsymbol{\Psi})
\end{aligned}
$$

where $\mathcal{N}(\boldsymbol{m}, \boldsymbol{V})$ is a Gaussian distribution with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{V}$.

1p **2a** Work out an equivalent expression for this model as a joint probability distribution over $\boldsymbol{x}_n$ and $\boldsymbol{z}_n$.

(a)
$$p(\boldsymbol{x}_n, \boldsymbol{z}_n) = \frac{\mathcal{N}(\boldsymbol{x}_n | \boldsymbol{W}\boldsymbol{z}_n, \boldsymbol{\epsilon}_n)}{\mathcal{N}(\boldsymbol{z}_n | 0, \boldsymbol{I})}$$

(b)
$$p(\boldsymbol{x}_n, \boldsymbol{z}_n) = \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{W}\boldsymbol{z}_n, \boldsymbol{\Psi})$$

(c)
$$p(\boldsymbol{x}_n, \boldsymbol{z}_n) = \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{W}\boldsymbol{z}_n, \boldsymbol{\epsilon}_n)\mathcal{N}(\boldsymbol{z}_n | 0, \boldsymbol{I})$$

(d)
$$p(\boldsymbol{x}_n, \boldsymbol{z}_n) = \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{W}\boldsymbol{z}_n, \boldsymbol{\Psi})\mathcal{N}(\boldsymbol{z}_n | 0, \boldsymbol{I})$$

2p **2b** Work out an expression for $p(\boldsymbol{x}_n)$.

(a) $p(\boldsymbol{x}_n) = \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{0}, \boldsymbol{W}\boldsymbol{W}^T + \boldsymbol{\Psi})$

(b) $p(\boldsymbol{x}_n) = \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{W}\boldsymbol{z}_n, \boldsymbol{W}\boldsymbol{W}^T + \boldsymbol{\Psi})$

(c) $p(\boldsymbol{x}_n) = \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{0}, \boldsymbol{W}^T\boldsymbol{W} + \boldsymbol{\Psi})$

(d) $p(\boldsymbol{x}_n) = \mathcal{N}(\boldsymbol{x}_n | \boldsymbol{W}\boldsymbol{z}_n, \boldsymbol{\epsilon}_n)$

2p **2c** This model is known as a "factor analysis" model and is commonly used to compress observations $\boldsymbol{x}_n$ into lower dimensional variables $\boldsymbol{z}_n$. Before we can make use of this model, we will need to train the parameters $\boldsymbol{W}$. Let's start by adding a prior

$$\boldsymbol{W} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) .$$

Consider an observed data set $X = \{\boldsymbol{x}_n | n = 1, 2, \ldots, N\}$. How would you train the parameters $\boldsymbol{W}$ for this application?

(a) Compute a posterior $p(\boldsymbol{W} | \{\boldsymbol{x}_n\}, \{\boldsymbol{z}_n\})$. This is easy because the joint is a Gaussian system so we can do this analytically with sum and product rules.

(b) Compute a posterior $p(\boldsymbol{W} | \{\boldsymbol{x}_n\}, \{\boldsymbol{z}_n\})$. This is hard because $\{\boldsymbol{z}_n\}$ is unobserved. Consider variational Bayesian approach.

(c) Compute a posterior $p(\boldsymbol{W} | \{\boldsymbol{x}_n\})$. This is hard because both $\{\boldsymbol{z}_n\}$ and $\boldsymbol{W}$ are unobserved. Consider a variational Bayesian approach.

(d) Compute a posterior $p(\boldsymbol{W} | \{\boldsymbol{x}_n\})$ through Bayes rule. This is easy because the joint is a Gaussian system so we can do this analytically with sum and product rules.

Question 3

You have a machine that measures property x, the "orangeness" of liquids. You wish to discriminate between C_1 = 'Fanta' and C_2 = 'Orangina'. It is known that

$$p(x|C_1) = \begin{cases} 1 & 1.0 \le x \le 2.0 \\ 0 & \text{otherwise} \end{cases}$$

$$p(x|C_2) = \begin{cases} 2 \cdot (x-1) & 1.0 \le x \le 2.0 \\ 0 & \text{otherwise} \end{cases}$$

The probability that x falls outside the interval $[1.0, 2.0]$ is zero. The prior class probabilities $p(C_1) = 0.6$ and $p(C_2) = 0.4$ are also known from experience.

1p **3a** We want to develop a Bayesian classifier. The discrimination boundary on the interval $x \in [1.0, 2.0]$ is given by

- a $1 = \frac{p(x|C_2)}{p(x|C_1)} \cdot \frac{p(C_1)}{p(C_2)} = \frac{1}{2(x-1)} \cdot \frac{0.4}{0.6} \quad \Rightarrow x = 5/3$
- b $1 = \frac{p(x|C_2)}{p(x|C_1)} = \frac{1}{2(x-1)} \quad \Rightarrow x = 3/2$
- c $1 = \frac{p(C_2|x)}{p(C_1|x)} = \frac{1 \cdot 0.6}{2(x-1) \cdot 0.4} \quad \Rightarrow x = 7/4$

2p **3b** Compute $p(C_1|x = 1.3)$.

- a $p(C_1|x = 1.3) = \frac{p(x=1.3|C_1)p(C_1)}{p(x=1.3|C_2)p(C_2)} = \frac{1 \cdot 0.6}{2(1.3-1) \cdot 0.4}$
- b $p(C_1|x = 1.3) = p(x = 1.3|C_1)p(C_1) = 1 \cdot 0.6$
- c $p(C_1|x = 1.3) = \frac{p(x=1.3|C_1)p(C_1)}{p(x=1.3)} = \frac{1 \cdot 0.6}{1 \cdot 0.6 + 2(1.3-1) \cdot 0.4}$

2p **3c** Let the discrimination boundary be given by $x = a$. Work out the total probability of a false classification:

- a $\int_{1.0}^{a} p(x|C_2)p(C_2)\mathrm{d}x + \int_{a}^{2} p(x|C_1)p(C_1)\mathrm{d}x$
- b $\int_{1.0}^{a} p(C_1|x)p(x)\mathrm{d}x + \int_{a}^{2} p(C_2|x)p(x)\mathrm{d}x$
- c $\int_{1.0}^{a} p(C_2|x)\mathrm{d}x + \int_{a}^{2} p(C_1|x)\mathrm{d}x$

This page is left blank intentionally