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

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

Bayesian machine learning and information processing (5SSD0)
5SSD0 Bayesian machine learning and information processing Q2 20-21 Part B Handwritten exam

Question 1

1p **1a** The Free Energy Principle (FEP) is a theory about biological self-organization, in particular about how brains develop through interactions with their environment. Which of the following statements is most consistent with FEP:

- (a) Our actions aim to reduce the complexity of our model of the environment.
- (b) Learning maximizes variational free energy
- (c) Perception aims to reduce the complexity of our model of the environment.
- (d) We act to fullfil our predictions about future sensory inputs.

1p **1b** Given is a data set $D = \{(x_n, y_n)\}_{n=1}^N$ where x_n holds a set of features and y_n is the class label for the n-th item. The discriminative approach to classification is based on a model proposal $p(y_n|x_n, \theta)$ with prior $p(\theta)$. After training this model on data D, the Bayesian class prediction $y_\bullet$ for a new input $x_\bullet$ is based on

- (a) $p(y_\bullet|x_\bullet, D) = \int p(y_\bullet|x_\bullet, \theta, D)\, \mathrm{d}\theta$
- (b) $p(y_\bullet|x_\bullet) = \int p(y_\bullet|x_\bullet, \theta)p(\theta)\mathrm{d}\theta$
- (c) $p(y_\bullet|x_\bullet, D) = \int p(y_\bullet|x_\bullet, \theta)p(\theta|D)\mathrm{d}\theta$
- (d) $p(y_\bullet|x_\bullet) = \int p(y_\bullet|x_\bullet, \theta)\mathrm{d}\theta$

1p **1c** In the Bayesian approach to model comparison, we rate the performance of model m_k for a given data set $D = \{x_n\}_{n=1}^N$ by its posterior probability $p(m_k|D)$, which can be evaluated by the following formula:

- (a) $p(m_k|D) = p(m_k) \int p(\theta|D, m_k)p(D|m_k)\mathrm{d}\theta$
- (b) $p(m_k|D) = p(m_k) \int p(D|\theta, m_k)p(\theta|m_k)\mathrm{d}\theta$
- (c) $p(m_k|D) = \sum_n p(m_k)p(D|m_k)$
- (d) $p(m_k|D) = \int p(D|\theta, m_k)p(\theta|m_k)\mathrm{d}\theta$

1p **1d** A dark bag contains five red balls and seven green ones. Balls are not returned to the bag after each draw. If you know that on the third draw the ball was a green one, what is now the probability of drawing a red ball on the first draw?

- (a) $4/12$
- (b) $5/11$
- (c) $5/12$
- (d) $4/11$

1p **1e** Which of the following expressions is a correct generative Gaussian Mixture Model for observations x_n and hidden one-hot coded cluster selection variables z_n?

- (a) $p(x_n, z_n) = \prod_{k=1}^{K} \left(\pi_k \cdot \mathcal{N}(x_n|\mu_k, \Sigma_k)\right)^{z_n}$
- (b) $p(x_n, z_n) = \prod_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_n|\mu_k, \Sigma_k)^{z_{nk}}$
- (c) $p(x_n, z_n) = \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(x_n|\mu_k, \Sigma_k)$
- (d) $p(x_n, z_n) = \prod_{k=1}^{K} \left(\pi_k \cdot \mathcal{N}(x_n|\mu_k, \Sigma_k)\right)^{z_{nk}}$

Question 2

A model m_1 is described by a single parameter θ, with $0 \le \theta \le 1$. The system can produce data $x \in \{0, 1, \ldots\}$.

The sampling distribution $p(x|\theta, m_1)$ and prior $p(\theta|m_1)$ are given by

$$p(x|\theta, m_1) = (1-\theta)\theta^x$$
$$p(\theta|m_1) = 6\theta(1-\theta)$$

1p **2a** Determine the posterior $p(\theta|x=4, m_1)$

- (a) $6\theta^4(1-\theta)^2$
- (b) $\frac{\int_0^1 \theta^5(1-\theta)^2 \mathrm{d}\theta}{\theta^5(1-\theta)^2}$
- (c) $\frac{\theta^5(1-\theta)^2}{\int_0^1 \theta^5(1-\theta)^2 \mathrm{d}\theta}$

1p **2b** Determine the probability $p(x=4|m_1)$.

- (a) $\int_0^1 \frac{(1-\theta)\theta^4}{6\theta(1-\theta)} \mathrm{d}\theta$
- (b) $\int_0^1 (1-\theta)\theta^4 \mathrm{d}\theta$
- (c) $\int_0^1 6\theta^5(1-\theta)^2 \mathrm{d}\theta$
- (d) $\int_0^1 \frac{6\theta(1-\theta)}{(1-\theta)\theta^4} \mathrm{d}\theta$

Consider a second model m_2 with the following sampling distribution and prior on $0 \le \theta \le 1$:

$$p(x|\theta, m_2) = (1-\theta)\theta^x$$
$$p(\theta|m_2) = 2\theta$$

The model priors are given by $p(m_1) = 2/3$ and $p(m_2) = 1/3$.

1p **2c** Determine the probability $p(x = 4|m_2)$.

(a) $\int_0^1 2(1-\theta)\theta^5 \mathrm{d}\theta$

(b) $\frac{1}{\int_0^1 2(1-\theta)\theta^5} \mathrm{d}\theta$

(d) $\int_0^1 \frac{(1-\theta)\theta^4}{2\theta} \mathrm{d}\theta$

1p **2d** Which of the two models has the largest Bayesian evidence after observing $x = 4$?

(a) m_1

(b) m_2

(c) same evidence for both models

1p **2e** Which of the two models has the largest posterior probability after observing $x = 4$?

(a) m_1

(b) m_2

(c) both models have the same posterior probability.

Question 3 (PP)

1p **3a** Suppose we have specified the following likelihood:

```
@RV X ~ Bernoulli(θ)
```

Which of the following is an appropriate prior specification:

(a) @RV $\theta \sim$ GaussianMeanVariance$(0, 1)$

(b) @RV $\theta \sim$ Beta$(1, 1)$

(c) @RV $\theta \sim$ Gamma$(1, 1)$

2p **3b** What is wrong with the following regression model specification?

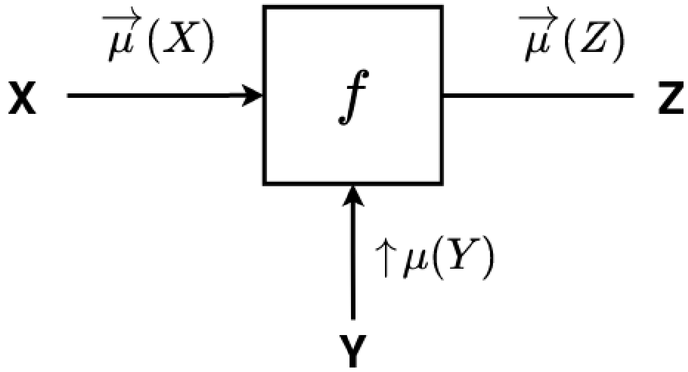
```
# Define covariates
@RV X

# Define weights prior
@RV θ ~ GaussianMeanVariance([0., 0.], [1. 0.; 0. 1.])

# Define likelihood
@RV Y ~ GaussianMeanVariance(dot(θ,X), [1. 0.; 0. 1.])
```

- (a) The weights prior should follow a Gamma distribution, not a Gaussian distribution.
- (b) The likelihood cannot have a dot product as a mean parameter.
- (c) The variance of the likelihood should be a scalar, not a matrix.
- (d) There is nothing wrong with this regression model specification.

2p **3c** Suppose we have the following factor graph:



where the factor node implements the function $f(X,Y,Z) = \delta(Z - XY)$.

Suppose the message from X is of the form $\overrightarrow{\mu}(X) = \mathcal{N}(X \mid 1,1)$ and the message from Y of the form $\uparrow\mu(Y) = \mathcal{N}(Y \mid 0,1)$. What is the form of the message $\overrightarrow{\mu}(Z)$?

- (a) $\mathcal{N}(Z \mid \frac{1}{2}, 2)$
- (b) $\mathcal{N}(Z \mid \frac{1}{2}, 1)$
- (c) $\mathcal{N}(Z \mid 0, 2)$
- (d) The message is not in the form of a Gaussian distribution.