Data Analytics and Machine Learning
Department of Informatics
Technical University of Munich

Eexam
Place student sticker here

Note:

- During the attendance check a sticker containing a unique code will be put on this exam.
- This code contains a unique number that associates this exam with your registration number.
- This number is printed both next to the code and to the signature field in the attendance check list.

Machine Learning for Graphs and Sequential Data

Exam:	IN2323 / Endterm	**Date:**	Friday 19th August, 2022
Examiner:	Prof. Dr. Stephan Günnemann	**Time:**	08:15 – 09:30

	P 1	P 2	P 3	P 4	P 5	P 6	P 7	P 8	P 9	P 10
I										

Working instructions

- This exam consists of **16 pages** with a total of **10 problems**.
  Please make sure now that you received a complete copy of the exam.
- The total amount of achievable credits in this exam is 86 credits.
- Detaching pages from the exam is prohibited.
- Allowed resources:
  - one A4 sheet of handwritten notes (two sides, not digitally written and printed).
- **No other material (e.g. books, cell phones, calculators) is allowed!**
- Physically turn off all electronic devices, put them into your bag and close the bag.
- There is scratch paper at the end of the exam (after problem 10).
- Write your answers only in the provided solution boxes or the scratch paper.
- If you solve a task on the scratch paper, clearly reference it in the main solution box.
- All sheets (including scratch paper) have to be returned at the end.
- **Only use a black or a blue pen (no pencils, red or greens pens!)**
- **For problems that say "Justify your answer" you only get points if you provide a valid explanation.**
- **For problems that say "Derive" you only get points if you provide a valid mathematical derivation.**
- **For problems that say "Prove" you only get points if you provide a valid mathematical proof.**
- If a problem does not say "Justify your answer", "Derive" or "Prove", it is sufficient to only provide the correct answer.

Left room from ________ to ________ / Early submission at ________

Problem 1 Normalizing flows (8 credits)

You are given the task of density estimation on $\mathbb{R}^2$ and plan on using normalizing flows. In the following we present some candidate transformations that will be used for **reverse parameterization.** For each of the transformations, state if it can be used to define a normalizing flow and justify your answers.

In all cases, the input is a vector $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$. We denote the output of the transformation with $\boldsymbol{z} \in \mathbb{R}^2$.

0
1
2

a)

$$\boldsymbol{A} = \boldsymbol{W}^T \boldsymbol{W}$$
$$\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x},$$

where $\boldsymbol{W} \in \mathbb{R}^{2\times 2}$.

0
1
2

b)

$$\boldsymbol{z} = \begin{bmatrix} x_1^2 & x_2^2 \end{bmatrix}^T$$

0
1
2

c)

$$z_1 = \boldsymbol{V}\mathrm{ReLU}(\boldsymbol{W}x_2 + \boldsymbol{b})$$
$$\boldsymbol{z} = \begin{bmatrix} z_1 & x_2 \end{bmatrix}^T,$$

where $\boldsymbol{W} \in \mathbb{R}^{h\times 1}$, $\boldsymbol{V} \in \mathbb{R}^{1\times h}$, $\boldsymbol{b} \in \mathbb{R}^h$ and ReLU is applied elementwise.

d)

0
1
2

$$\boldsymbol{z} = \boldsymbol{a} \odot \boldsymbol{x} + \boldsymbol{b},$$

where $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^2$ and $\odot$ is the elementwise product.

Problem 2 Variational inference (10 credits)

0
1
2
3
4
5
6
7
8
9
10

Suppose we are given a latent variable model for a sequence of observations $x_1, \dots, x_N \in \{0, 1\}$ and latent variables $z_1, \dots, z_N \in [0, 1]$ with

$$p(z_1, \dots, z_N) = \prod_{n=1}^{N} \text{Beta}(z_n \mid \alpha, \beta) = \prod_{n=1}^{N} \frac{1}{B(\alpha, \beta)} z_n^{\alpha-1}(1 - z_n)^{\beta-1}$$

$$p(x_1, \dots, x_N | z_1, \dots, z_N) = \prod_{n=1}^{N} \text{Bern}(x_n \mid z_n) = \prod_{n=1}^{N} z_n^{x_n}(1 - z_n)^{1-x_n}$$

with parameters $\alpha, \beta > 0$ and normalizing constant $B(\alpha, \beta)$. We define the variational distribution

$$q(z_1, \dots, z_N) = \prod_{n=1}^{N} \text{Beta}(z_n \mid \gamma, \delta) = \prod_{n=1}^{N} \frac{1}{B(\gamma, \delta)} z_n^{\gamma-1}(1 - z_n)^{\delta-1}$$

with parameters $\gamma, \delta > 0$.
Assume that α, β are known and fixed. Prove or disprove the following statement:
There **exist** observations $x_1, \dots, x_N \in \{0, 1\}$ and values of $\gamma, \delta > 0$ such that the ELBO is tight, i.e. $\exists x_1, \dots, x_N, \exists \gamma, \delta : \log(p(x_1, \dots, x_N)) = \mathcal{L}((\alpha, \beta), (\gamma, \delta))$.

Problem 3 Robustness (9 credits)

In the lecture, we have derived a convex relaxation for the ReLU activation function. Now, we want to generalize this result to the flexible ReLU (FReLU) activation function

0
1
2
3
4
5
6
7
8
9

$$FReLU(x) = \begin{cases} x + b & \text{if } x > 0 \\ b & \text{if } x \leq 0 \end{cases}$$

with variable input $x \in \mathbb{R}$ and **constant parameter** $b \in \mathbb{R}$.

Let $y \in \mathbb{R}$ be the variable we use to model the function's output. Now, given input bounds $l, u \in \mathbb{R}$ with $l \leq x \leq u$, provide a set of **linear constraints** corresponding to the convex hull of $\{\begin{bmatrix} x & FRelu(x) \end{bmatrix}^T \mid l \leq x \leq u\}$.

Hint: You will have to make a case distinction to account for different ranges of l and u.

Problem 4 **Autoregressive models (8 credits)**

0
1
2
3
4
5

a)
An autoregressive process of order p, AR(p), is defined as:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \, ,$$

with independently distributed noise variables $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.
Provided that the AR(p) is stationary, derive its first moment $\mathbb{E}[X_t]$ as a function of c and φ_i.

0
1
2
3

b) Let us define a process X_t as

$$X_t = \sin^2\left(-\frac{\pi}{2}t\right) + \frac{2}{3}X_{t-1} + \varepsilon_t$$

with independently distributed noise variables $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.
Decide if the process X_t is stationary. Justify your answer.

Problem 5 Hidden Markov Models (9 credits)

Consider a hidden Markov model with 2 states $\{1, 2\}$ and 4 possible observations $\{\mathrm{c}, \mathrm{e}, \mathrm{i}, \mathrm{n}\}$. The initial distribution π, transition probabilities $\boldsymbol{A}$ and emission probabilities $\boldsymbol{B}$ are

$$\pi = \begin{array}{c} 1 \\ 2 \end{array} \begin{pmatrix} 2/5 \\ 3/5 \end{pmatrix} \qquad \boldsymbol{A} = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} 1/3 & 2/3 \\ 3/5 & 2/5 \end{pmatrix} \end{array} \qquad \boldsymbol{B} = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cccc} \mathrm{c} & \mathrm{e} & \mathrm{i} & \mathrm{n} \end{array} \\ \begin{pmatrix} 2/5 & 1/5 & 1/5 & 1/5 \\ 0 & 1/5 & 3/5 & 1/5 \end{pmatrix} \end{array},$$

where $\boldsymbol{A}_{ij}$ specifies the probability of transitioning from state i to state j.

a) We have observed the sequence $X_{1:3} = [\mathrm{nic}]$. What is the most likely latent state Z_3 given these observations? Justify your answer. What is this type of inference called?

0 1 2 3

b) The full observed sequence is $X_{1:4} = [\mathrm{nice}]$. What is the most likely latent state sequence $Z_{1:4}$ given these observations? Justify your answer.

0 1 2 3 4 5 6

Problem 6 **Temporal Point Processes (10 credits)**

0 1 2 3 4 5 6 7 8 9 10

Assume that we use a Hawkes process to model a discrete event sequence $\{t_1, \dots, t_N\}$ with $t_i \in [0, T]$. Further assume that (like in the lecture) we use an exponential triggering kernel, i.e. $k_\omega(t - t_i) = \exp(-\omega(t - t_i))$. Prove that the log-likelihood-function of the process is

$$\log p_\theta(\{t_1, \dots, t_N\}) = \sum_{i=1}^{N} \log \left(\mu + \alpha \sum_{j<i} \exp(-\omega(t_i - t_j)) \right) - \mu T + \frac{\alpha}{\omega} \sum_{i=1}^{N} (\exp(-\omega(T - t_i)) - 1)$$

Problem 7 Graphs – Generative Models (8 credits)

Let $\boldsymbol{A} \in \{0,1\}^{N \times N}$ be the adjacency matrix of a graph generated by a stochastic block model with $\boldsymbol{\pi} = \begin{bmatrix} a & 1-a \end{bmatrix}^T$, $\boldsymbol{\eta} = \begin{bmatrix} p & q \\ q & p \end{bmatrix}$ and parameters $a, p, q \in [0,1]$. Let $\deg(n) = \sum_{j=1}^{N} A_{n,j}$ be the degree of node n.

Derive the expected degree $\mathbb{E}[\deg(n)]$ of an arbitrary node n.

0
1
2
3
4
5
6
7
8

Problem 8 Graphs – Clustering (10 credits)

Let $\boldsymbol{A} \in \{0,1\}^{N \times N}$ be the adjacency matrix of an undirected graph (i.e. symmetric adjacency matrix) generated by a stochastic block model with $\boldsymbol{\pi} = \begin{bmatrix} a & 1-a \end{bmatrix}^T$, $\boldsymbol{\eta} = \begin{bmatrix} p & q \\ q & p \end{bmatrix}$ and parameters $a, p, q \in [0,1]$.

0 1 2

a) Assume that $p, q, a \in [0,1]$ are known and fixed. Can $\Pr(\boldsymbol{z} \mid \boldsymbol{A}, \boldsymbol{\eta}, \boldsymbol{\pi})$, the probability mass function of community assignments $\boldsymbol{z}$ given $\boldsymbol{A}$, be evaluated in polynomial time? That is, can it be evaluated in $\mathcal{O}(N^c)$, where N is the number of nodes and $c \in \mathbb{R}_+$? Justify your answer.

0 1 2 3 4 5 6 7 8

b) Now assume that $p = 0$. Further assume that $\boldsymbol{A}$ is a connected graph (i.e. each pair of nodes (i, j) is connected by a path). Propose a procedure for finding the most likely community assignment, i.e.

$$\max_{\boldsymbol{z} \in \{0,1\}^N} \Pr(\boldsymbol{z} \mid \boldsymbol{A}, \boldsymbol{\eta}, \boldsymbol{\pi})$$

in polynomial time $\mathcal{O}(N^c)$. Justify your answers.

Problem 9 Limitations of Graph Neural Networks (6 credits)

a) What is oversmoothing and what causes it?

0
1
2
3

b) The Personalized Propagation of Neural Predictions (PPNP) architecture is designed to overcome the problem of oversmoothing. Briefly explain its two key building blocks.

0
1
2
3

Problem 10 Page Rank (8 credits)

Recall the spam farm discussed in our exercise. It consists of the spammer's own pages S_{own} with target page t and k supporting pages, as well as links from the accessible pages S_{acc} to the target page. **Different from the exercise, every page within S_{own} has a link to every other page within S_{own}** (see Fig. 10.1).

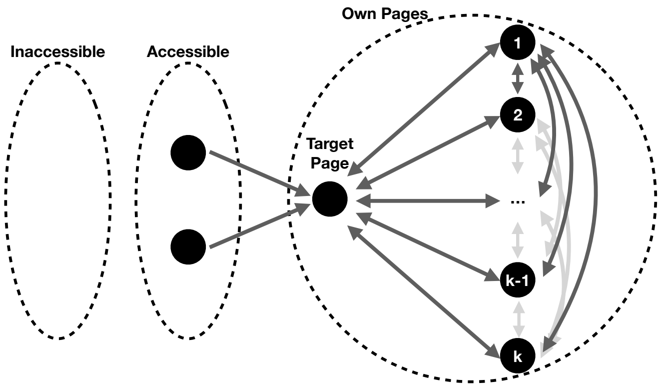

Figure 10.1

Let n be the total number of pages on the web, E the set of all edges, r_p the PageRank score of a page p and d_p the degree of a page p. Let $x_{\text{acc}} = \sum_{p \in S_{\text{acc}}, (p,t) \in E} \frac{r_p}{d_p}$ be the amount of PageRank contributed by the accessible pages. We are using PageRank with teleports, where $(1 - \beta)$ is the teleport probability.

0 1 2 3

a) Derive the PageRank r_s of a supporting page r_s as a function of β, r_t, k, n.

b) Derive the PageRank r_t of the target page as a function of x_{acc}, k, β, n. You do not have to simplify.

0
1
2
3

c) How can the spammer modify the edges of the k supporting pages to increase the PageRank score r_t of the target page? Justify your answer.

0
1
2

Additional space for solutions–clearly mark the (sub)problem your answers are related to and strike out invalid solutions.