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 $x_1, \ldots, x_N \in \{0, 1\}$ and latent variables $z_1, \ldots, z_N \in [0, 1]$ with

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

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

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

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

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

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

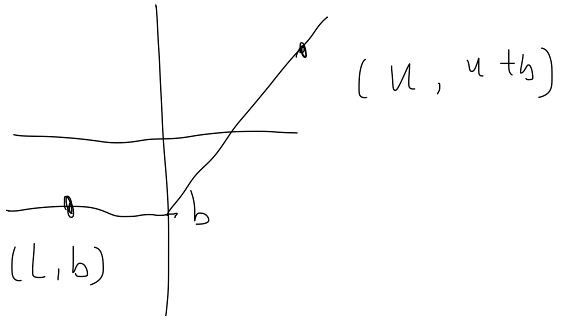
0
1
2
3
4
5
6
7
8
9

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

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 $\{\begin{bmatrix} x & FRelu(x) \end{bmatrix}^T | l \leq x \leq u\}$.

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



$l, u \le 0$

$y = b$

$l, u > 0$

$y = x + b$

$l < 0 < u$

$y \ge b$

$y \ge x + b$

$$y = ax + b$$

$$b = l \cdot a + b$$

$$u + b = u \cdot a + b$$

$$u = (u - l) \cdot a$$

$$a = \frac{u}{u - l}$$

$$b = \frac{ul}{u - l} + b$$

$$b = \frac{ub - lb - ul}{u - l}$$

$$y \le \frac{u}{u - l} x + \frac{ub - lb - ul}{u - l}$$

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.

$$E(X_t) = C + \sum_{i=1}^{p} \varphi_i E(X_{t-i}) + \underbrace{E(\varepsilon_t)}_{0}.$$

$$E(X_t) = \frac{C}{1 - \sum_{i=1}^{p} \varphi_i}$$

---

assue stationary

$$\mu = \sin^2\left(-\frac{\pi}{2}t\right) + \frac{2}{3}\mu$$

$$\frac{1}{3}\mu = \sin^2\left(-\frac{\pi}{2}t\right)$$

chaη by t

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

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

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

$\alpha_1 = \pi \odot B$

$$= \begin{pmatrix} \frac{2}{5} \\ \frac{3}{5} \end{pmatrix} \odot \begin{pmatrix} \frac{1}{5} \\ \frac{1}{5} \end{pmatrix} = \begin{pmatrix} \frac{2}{25} \\ \frac{3}{25} \end{pmatrix}$$

$\alpha_2 = B \odot (A^T \alpha_1)$

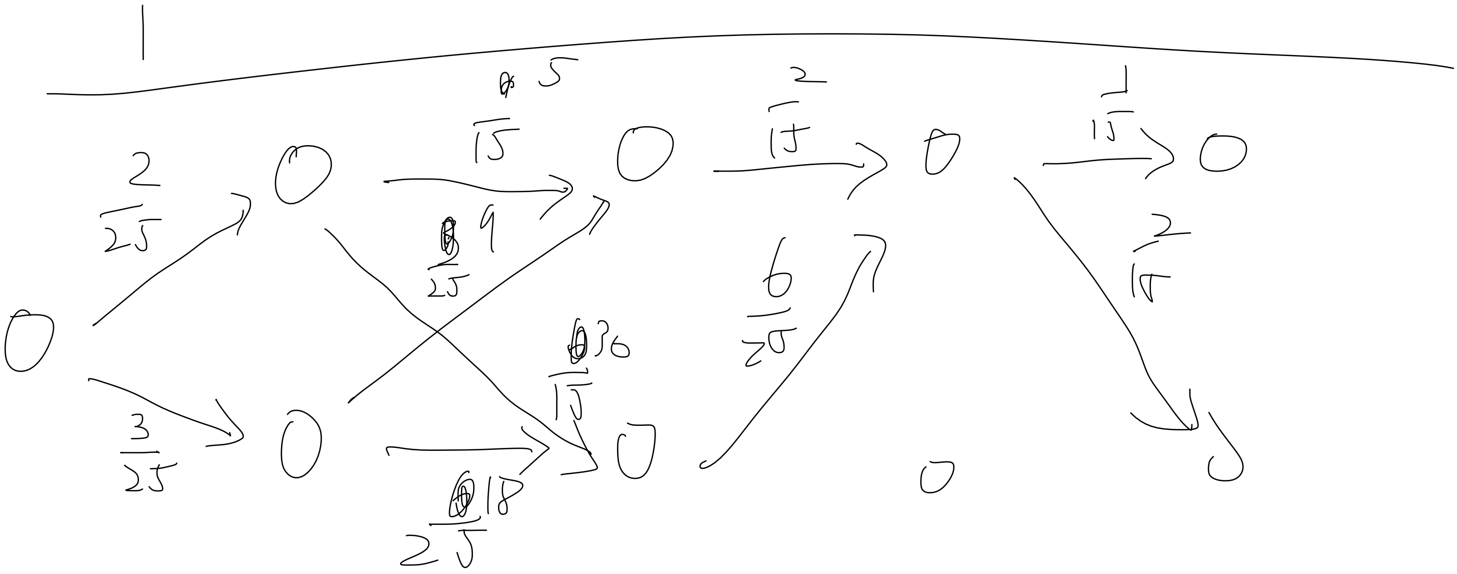
$$= \begin{pmatrix} \frac{1}{5} \\ \frac{3}{5} \end{pmatrix} \odot \begin{pmatrix} \frac{1}{3} & \frac{3}{5} \\ \frac{2}{3} & \frac{2}{5} \end{pmatrix} \begin{pmatrix} \frac{2}{25} \\ \frac{3}{25} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{5} \\ \frac{3}{5} \end{pmatrix} \odot \begin{pmatrix} \frac{2}{25}\left(\frac{4}{15}\right) \\ \frac{3}{25}\left(\frac{10}{15}\right) \end{pmatrix}$$

$$= \begin{matrix} \frac{2}{125} \cdot \frac{14}{15} \\ \frac{9}{125} \cdot \frac{16}{15} \end{matrix}$$

$$\alpha_3 = \begin{pmatrix} \frac{2}{5} \\ 0 \end{pmatrix} \odot \Big($$

1



1 2 1 2

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

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

$$\lambda^*(t) = \mu + \alpha \sum_{t_j \in H(t)} \underline{K_w(t-t_j)}$$

$\hookrightarrow$ Triggering kernel $= \exp(-w(t-t_i))$

$$P = \sum \log \lambda^*(t) \quad - \int_0^T \lambda^*(u)\, du$$

$$= \sum_{i=1}^{N} (\mu + \alpha \sum_{j<1} \exp(-w(t-t_i)) \quad -$$

$$\int_0^T \underline{\mu} + \underline{\alpha \sum \exp(-w(u-t_i))} \quad du.$$

$$\cancel{\mu u|_0^T + \frac{\alpha}{-w} \sum \exp(-w(T-t_i))}$$

$\lambda(u)$ change after each event

$$\int_0^{t_1} \underline{\mu}\, du + \sum_{i=1}^{N} \int_{t_i}^{t_{i+1}} \mu \quad + \alpha \sum \exp(-w(t-t_i)) \cdot d\mu$$

看 summary

$$\Downarrow$$

$$\mu T$$

$$\alpha \sum_{i=1}^{N} \sum_{j<i} \int_{t_i}^{t_{i+1}} \exp(-w(t-t_j)) d\mu$$

$$-\frac{\alpha}{w} \sum_{i=1}^{N} \sum_{j<i} \left[\exp(-w(t_{i+1}-t_j)) - \exp(-w(t_i-t_j))\right]$$

$j=i$

$$-\frac{\alpha}{w}\left[\exp(-w(T-t_i)) - 1\right]$$

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

$$E[\deg(n)] = \sum_{j=1}^{N} E[A_{n,j}]$$

$$\sum_{j=1}^{N-1} E(A_{N,j}) = \sum_{j=1}^{N-1} Pr[A_{n,j} = 1]$$

$$= p(A_{N,j} = 1) =$$

$$p(A_{N,j} | z_i, z_j \eta) \cdot \underline{p(z_i = z_j = 1)}$$

$$+ p(A_{N,j} | z_i, z_j \eta)\, p(z_i = z_j = 2)$$

$$+ p(A_{N,j} | z_i, z_j \eta)$$

$$p(z_i = 1, z_j = 2)$$

$$+ p(A_{N,j} | z_i, z_j \eta)$$

$$\text{Bern}(A_{ij}) = p * a \cdot a +$$

$$p * (1-a)^2 +$$

$$2 * q * a(1-a)$$

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.

$$p(Z|A,\eta,\lambda) = \frac{p(A|Z,\cdots)\cdot p(Z|\cdots)}{p(A|\cdots)}$$

For each Node inde all converted node M.

For all nodes chock no more the N^N

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

the nodes in infinit graph step
tend to be a stationary distribution

embedding vectors of each node
$\Rightarrow$ same constant vector.

Network architecture

Influence of all node follow pagerank

depend on whole graph not local
neighbour

---

seperate feature Translation and propagation.
$\Rightarrow$ apply NN on each node
before propagation

apply personalised teleporting probability.

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).

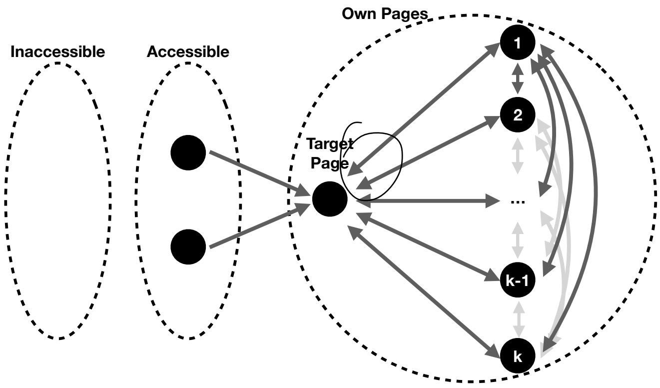


Figure 10.1

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.

$$r_s = \sum^{k-1} \beta \frac{r_s}{k} + \beta \frac{r_t}{k} + (1-\beta) \cdot \frac{1}{n}$$

$$= \beta \frac{k-1}{k} r_s + \beta \frac{r_t}{k} + (1-\beta) \frac{1}{n}$$

$$r_s \left(1 - \beta \cdot \frac{k-1}{k}\right) = \beta \cdot \frac{r_t}{k} + (1-\beta) \frac{1}{n}$$

$$r_s = \frac{\beta \cdot r_t}{k - \beta \cdot (k-1)} + \frac{k(1-\beta)}{n \cdot (k - \beta(k-1))}$$

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

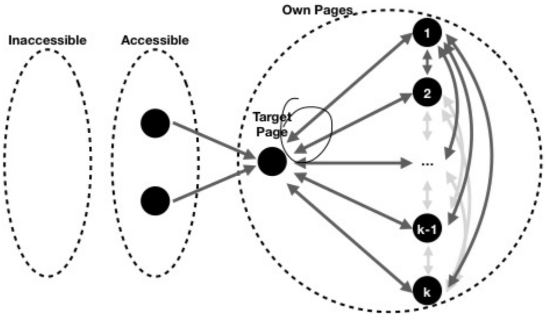


Figure 10.1

$$r_s = \frac{\beta \cdot r_t}{k - \beta \cdot (k-1)} + \frac{k(1-\beta)}{n \cdot (k - \beta(k-1))}$$

$$r_t = \beta x_{acc} + \beta \sum_{i=1}^{k} \frac{r_i}{k} + (1-\beta) \cdot \frac{1}{n}$$

$$= \beta x_{acc} + \frac{\beta^2 r_t}{k - \beta(k-1)} + \frac{\beta k (1-\beta)}{n \cdot (k - \beta(k-1))} + (1-\beta)\frac{1}{n}$$

$$r_t = \left(1 - \frac{\beta^2}{k - \beta(k-1)} + \ldots\right.$$

---

decreasing number of edges

remove edge

better support nodes

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