Data Analytics and Machine Learning Group
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 (Problem sheet)

Graded Exercise:	IN2323 / Endterm	**Date:**	Friday 30th July, 2021
Examiner:	Prof. Dr. Stephan Günnemann	**Time:**	11:30 – 12:45

Working instructions

- **DO NOT SUBMIT THIS SHEET! ONLY SUBMIT YOUR PERSONALIZED ANSWER SHEET THAT IS DISTRIBUTED THROUGH TUMEXAM!**
- Make sure that you solve the version of the problem stated on your personalized answer sheet (e.g., Problem 1 (Version B), Problem 2 (Version A), etc.)

Problem 1: Normalizing Flows (Version A)

We use the following normalizing flow model to specify a density $p_2(\boldsymbol{x})$ over $\mathbb{R}^2$.
The base density $p_1(\boldsymbol{z})$ is equal to the $\mathrm{Uniform}([0,2]^2)$ distribution, that is

$$p_1(\boldsymbol{z}) = \begin{cases} \frac{1}{4} & \text{if } z_1 \in [0,2] \text{ and } z_2 \in [0,2], \\ 0 & \text{else.} \end{cases}$$

The model is specified by the following forward transformation $f\colon \mathbb{R}^2 \to \mathbb{R}^2$

$$\boldsymbol{x} = f(\boldsymbol{z}) = \boldsymbol{A}\boldsymbol{z} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/4 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

0
1
2
3
4
5
6

Compute the density $p_2(\boldsymbol{x})$ at the point $\boldsymbol{x}^{(0)} = (1/2, 1/8)^T$. Justify your answer.

Problem 1: Normalizing Flows (Version B)

We use the following normalizing flow model to specify a density $p_2(\boldsymbol{x})$ over $\mathbb{R}^2$.
The base density $p_1(\boldsymbol{z})$ is equal to the $\text{Uniform}([0,2]^2)$ distribution, that is

$$p_1(\boldsymbol{z}) = \begin{cases} \frac{1}{4} & \text{if } z_1 \in [0,2] \text{ and } z_2 \in [0,2], \\ 0 & \text{else.} \end{cases}$$

The model is specified by the following forward transformation $f \colon \mathbb{R}^2 \to \mathbb{R}^2$

$$\boldsymbol{x} = f(\boldsymbol{z}) = \boldsymbol{A}\boldsymbol{z} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/12 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Compute the density $p_2(\boldsymbol{x})$ at the point $\boldsymbol{x}^{(0)} = (1/2, 1/8)^T$. Justify your answer.

0 1 2 3 4 5 6

Problem 1: Normalizing Flows (Version C)

We use the following normalizing flow model to specify a density $p_2(\boldsymbol{x})$ over $\mathbb{R}^2$.
The base density $p_1(\boldsymbol{z})$ is equal to the $\text{Uniform}([0,2]^2)$ distribution, that is

$$p_1(\boldsymbol{z}) = \begin{cases} \frac{1}{4} & \text{if } z_1 \in [0,2] \text{ and } z_2 \in [0,2], \\ 0 & \text{else.} \end{cases}$$

The model is specified by the following forward transformation $f\colon \mathbb{R}^2 \to \mathbb{R}^2$

$$\boldsymbol{x} = f(\boldsymbol{z}) = \boldsymbol{A}\boldsymbol{z} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/12 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

0 1 2 3 4 5 6

Compute the density $p_2(\boldsymbol{x})$ at the point $\boldsymbol{x}^{(0)} = (1/2, 1/8)^T$. Justify your answer.

Problem 1: Normalizing Flows (Version D)

We use the following normalizing flow model to specify a density $p_2(\boldsymbol{x})$ over $\mathbb{R}^2$.
The base density $p_1(\boldsymbol{z})$ is equal to the Uniform$([0,2]^2)$ distribution, that is

$$p_1(\boldsymbol{z}) = \begin{cases} \frac{1}{4} & \text{if } z_1 \in [0,2] \text{ and } z_2 \in [0,2], \\ 0 & \text{else.} \end{cases}$$

The model is specified by the following forward transformation $f \colon \mathbb{R}^2 \to \mathbb{R}^2$

$$\boldsymbol{x} = f(\boldsymbol{z}) = \boldsymbol{A}\boldsymbol{z} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/10 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

Compute the density $p_2(\boldsymbol{x})$ at the point $\boldsymbol{x}^{(0)} = (1/2, 1/8)^T$. Justify your answer.

0
1
2
3
4
5
6

Problem 2: Variational Inference (Version A)

We would like to draw samples from a logistic distribution with reparametrization.
The cumulative distribution function (CDF) of the logistic distribution with scale parameter ϕ is defined as

$$F_x(a) = \Pr(x \leq a) = \frac{1}{1 + \exp(-\phi a)}.$$

We have access to an algorithm that produces samples u from the uniform distribution on $[0, 1]$, that is

$$b(u) = \begin{cases} 1 & \text{if } u \in [0, 1], \\ 0 & \text{else.} \end{cases}$$

0 1 2 3 4 5 6

Derive a deterministic transformation T_ϕ that converts a sample $u \sim b(u)$ into a sample x from the logistic distribution.

Problem 2: Variational Inference (Version B)

We would like to draw samples from a logistic distribution with reparametrization.
The cumulative distribution function (CDF) of the logistic distribution with scale parameter ϕ is defined as

$$F_x(a) = \Pr(x \leq a) = \frac{1}{1 + \exp(-a + \phi)}.$$

We have access to an algorithm that produces samples u from the uniform distribution on $[0, 1]$, that is

$$b(u) = \begin{cases} 1 & \text{if } u \in [0, 1], \\ 0 & \text{else.} \end{cases}$$

Derive a deterministic transformation T_ϕ that converts a sample $u \sim b(u)$ into a sample x from the logistic distribution.

☐ 0
☐ 1
☐ 2
☐ 3
☐ 4
☐ 5
☐ 6

Problem 2: Variational Inference (Version C)

We would like to draw samples from a logistic distribution with reparametrization.
The cumulative distribution function (CDF) of the logistic distribution with scale parameter ϕ is defined as

$$F_x(a) = \Pr(x \leq a) = \frac{1}{1 + \exp(-a/\phi)}.$$

We have access to an algorithm that produces samples u from the uniform distribution on $[0, 1]$, that is

$$b(u) = \begin{cases} 1 & \text{if } u \in [0, 1], \\ 0 & \text{else.} \end{cases}$$

0 Derive a deterministic transformation T_ϕ that converts a sample $u \sim b(u)$ into a sample x from the logistic
1 distribution.
2
3
4
5
6

Problem 2: Variational Inference (Version D)

We would like to draw samples from a logistic distribution with reparametrization.
The cumulative distribution function (CDF) of the logistic distribution with scale parameter ϕ is defined as

$$F_x(a) = \Pr(x \leq a) = \frac{1}{1 + \exp(-a - \phi)}.$$

We have access to an algorithm that produces samples u from the uniform distribution on $[0, 1]$, that is

$$b(u) = \begin{cases} 1 & \text{if } u \in [0, 1], \\ 0 & \text{else.} \end{cases}$$

Derive a deterministic transformation T_ϕ that converts a sample $u \sim b(u)$ into a sample x from the logistic distribution.

0
1
2
3
4
5
6

Problem 3: Deep Generative Models (Version A)

Suppose $f\colon \mathbb{R}^N \to \mathbb{R}^M$ is a fully-connected neural network, where $N < M$.

0 1 2

a) Can we use f as the generator in a generative adversarial network? Justify your answer.

0 1 2

b) Can we use f as the transformation to define a normalizing flow model? Justify your answer.

0 1 2

c) Can we use f as the decoder in a variational autoencoder model? Justify your answer.

Problem 3: Deep Generative Models (Version B)

Suppose $f\colon \mathbb{R}^N \to \mathbb{R}^M$ is a fully-connected neural network, where $N < M$.

a) Can we use f as the decoder in a variational autoencoder model? Justify your answer.

0
1
2

b) Can we use f as the transformation to define a normalizing flow model? Justify your answer.

0
1
2

c) Can we use f as the generator in a generative adversarial network? Justify your answer.

0
1
2

Problem 3: Deep Generative Models (Version C)

Suppose $f\colon \mathbb{R}^N \to \mathbb{R}^M$ is a fully-connected neural network, where $N < M$.

0 ☐
1 ☐
2 ☐

a) Can we use f as the transformation to define a normalizing flow model? Justify your answer.

0 ☐
1 ☐
2 ☐

b) Can we use f as the generator in a generative adversarial network? Justify your answer.

0 ☐
1 ☐
2 ☐

c) Can we use f as the decoder in a variational autoencoder model? Justify your answer.

Problem 3: Deep Generative Models (Version D)

Suppose $f\colon \mathbb{R}^N \to \mathbb{R}^M$ is a fully-connected neural network, where $N < M$.

a) Can we use f as the generator in a generative adversarial network? Justify your answer.

0
1
2

b) Can we use f as the decoder in a variational autoencoder model? Justify your answer.

0
1
2

c) Can we use f as the transformation to define a normalizing flow model? Justify your answer.

0
1
2

Problem 4: Robustness (Version A)

0 1 2 3 4 5 6 7 8

Assume we are interested in robustness certification for a model classifying data from $\mathbb{R}^D$. So far, we have only considered norm-bound perturbation models that — for the l_2 norm — can be expressed via the constraints

$$\begin{aligned} ||\tilde{\boldsymbol{x}} - \boldsymbol{x}||_2 &\leq \epsilon, \\ \tilde{\boldsymbol{x}} &\in \mathbb{R}^D, \end{aligned} \tag{4.1}$$

with adversarial budget $\epsilon \geq 0$, perturbed input $\tilde{\boldsymbol{x}}$ and unperturbed input $\boldsymbol{x} \in \mathbb{R}^D$.

Now we want to model *sparse* norm-bound perturbations, in which the adversary can perturb at most $\eta \in \mathbb{N}$ vector entries. For this, extend Equation (4.1) with additional **linear constraints**. You may introduce at most $\mathcal{O}(D)$ constraints and $\mathcal{O}(D)$ variables. You are allowed to use integer-valued variables.

More formally, the modeled perturbation set should be

$$\mathcal{P}(x) = \left\{ \tilde{\boldsymbol{x}} \in \mathbb{R}^D \,\middle|\, ||\tilde{\boldsymbol{x}} - \boldsymbol{x}||_2 \leq \epsilon \wedge \underline{\sum_{d=0}^{D-1} \mathbb{1}\left(\tilde{x}_d \neq x_d\right) \leq \eta} \right\},$$

计数

where $\mathbb{1}(\cdot)$ is the indicator function.

assume $q \in \{0, 1\}^D$ $\quad \sum q_d \leq \eta$

$\hat{x} - x \leq q_d$

$\hat{x} - x \geq q_d$

$q_d = 0 \Rightarrow \hat{x} = x$

$q_d = 1 \Rightarrow \hat{x} - x = 1$

Problem 4: Robustness (Version B)

0 1 2 3 4 5 6 7 8

Assume we are interested in robustness certification for a model classifying data from $\mathbb{R}^D$. So far, we have only considered norm-bound perturbation models that — for the l_2 norm — can be expressed via the constraints

$$\begin{aligned} ||\tilde{\boldsymbol{x}} - \boldsymbol{x}||_2 &\leq \epsilon, \\ \tilde{\boldsymbol{x}} &\in \mathbb{R}^D, \end{aligned} \tag{4.1}$$

with adversarial budget $\epsilon \geq 0$, perturbed input $\tilde{\boldsymbol{x}}$ and unperturbed input $\boldsymbol{x} \in \mathbb{R}^D$.

Now we want to model *sparse* norm-bound perturbations, in which the adversary can perturb at most $\eta \in \mathbb{N}$ vector entries. For this, extend Equation (4.1) with additional **linear constraints**. You may introduce at most $\mathcal{O}(D)$ constraints and $\mathcal{O}(D)$ variables. You are allowed to use integer-valued variables.

More formally, the modeled perturbation set should be

$$\mathcal{P}(x) = \left\{ \tilde{\boldsymbol{x}} \in \mathbb{R}^D \,\middle|\, ||\tilde{\boldsymbol{x}} - \boldsymbol{x}||_2 \leq \epsilon \wedge \sum_{d=0}^{D-1} \mathbb{1}\left(\tilde{x}_d \neq x_d\right) \leq \eta \right\},$$

where $\mathbb{1}(\cdot)$ is the indicator function.

Problem 4: Robustness (Version C)

0 1 2 3 4 5 6 7 8

Assume we are interested in robustness certification for a model classifying data from $\mathbb{R}^D$. So far, we have only considered norm-bound perturbation models that — for the l_2 norm — can be expressed via the constraints

$$\begin{aligned} ||\tilde{\boldsymbol{x}} - \boldsymbol{x}||_2 &\leq \epsilon, \\ \tilde{\boldsymbol{x}} &\in \mathbb{R}^D, \end{aligned} \tag{4.1}$$

with adversarial budget $\epsilon \geq 0$, perturbed input $\tilde{\boldsymbol{x}}$ and unperturbed input $\boldsymbol{x} \in \mathbb{R}^D$.

Now we want to model *sparse* norm-bound perturbations, in which the adversary can perturb at most $\eta \in \mathbb{N}$ vector entries. For this, extend Equation (4.1) with additional **linear constraints**. You may introduce at most $\mathcal{O}(D)$ constraints and $\mathcal{O}(D)$ variables. You are allowed to use integer-valued variables.

More formally, the modeled perturbation set should be

$$\mathcal{P}(x) = \left\{ \tilde{\boldsymbol{x}} \in \mathbb{R}^D \,\middle|\, ||\tilde{\boldsymbol{x}} - \boldsymbol{x}||_2 \leq \epsilon \wedge \sum_{d=0}^{D-1} \mathbb{1}\left(\tilde{x}_d \neq x_d\right) \leq \eta \right\},$$

where $\mathbb{1}(\cdot)$ is the indicator function.

Problem 4: Robustness (Version D)

Assume we are interested in robustness certification for a model classifying data from $\mathbb{R}^D$. So far, we have only considered norm-bound perturbation models that — for the l_2 norm — can be expressed via the constraints

$$\begin{aligned} ||\tilde{\boldsymbol{x}} - \boldsymbol{x}||_2 &\leq \epsilon, \\ \tilde{\boldsymbol{x}} &\in \mathbb{R}^D, \end{aligned} \tag{4.1}$$

with adversarial budget $\epsilon \geq 0$, perturbed input $\tilde{\boldsymbol{x}}$ and unperturbed input $\boldsymbol{x} \in \mathbb{R}^D$.

Now we want to model *sparse* norm-bound perturbations, in which the adversary can perturb at most $\eta \in \mathbb{N}$ vector entries. For this, extend Equation (4.1) with additional **linear constraints**. You may introduce at most $\mathcal{O}(D)$ constraints and $\mathcal{O}(D)$ variables. You are allowed to use integer-valued variables.

More formally, the modeled perturbation set should be

$$\mathcal{P}(x) = \left\{ \tilde{\boldsymbol{x}} \in \mathbb{R}^D \,\middle|\, ||\tilde{\boldsymbol{x}} - \boldsymbol{x}||_2 \leq \epsilon \wedge \sum_{d=0}^{D-1} \mathbb{1}\left(\tilde{x}_d \neq x_d\right) \leq \eta \right\},$$

where $\mathbb{1}(\cdot)$ is the indicator function.

Problem 5: Autoregressive Models (Version A)

Consider a sequential process described by the following autoregressive model.

$$X_t = X_{t-1} + \varepsilon_t$$

The noise variables ε_t are independently $\mathcal{N}(0, 1)$ distributed and we condition the process on $X_0 = 0$.

0 1 2

a) Is this autoregressive process stationary? Justify your answer.

$X_t = L X_t + \varepsilon_t$

$(1-L) = 0$

$L = 1 \qquad |L| \not> 1 \qquad$ unstationary

0 1 2

b) Now we define a second process Y_t as following

$$Y_t = X_t - X_{t-1}.$$

Is the process Y_t stationary? Justify your answer.

$Y_t + \cancel{X_{t-1}} = X_t = \cancel{X_{t-1}} + \varepsilon_t$

$Y_t = \varepsilon_t$

$E(Y_t) = 0 \qquad$ Yes

Problem 5: Autoregressive Models (Version B)

Consider a sequential process described by the following autoregressive model.

$$X_t = X_{t-1} + \varepsilon_t$$

The noise variables ε_t are independently $\mathcal{N}(0, 1)$ distributed and we condition the process on $X_0 = 0$.

a) Is this autoregressive process stationary? Justify your answer.

0
1
2

b) Now we define a second process Y_t as following

0
1
2

$$Y_t = X_t - X_{t-1}.$$

Is the process Y_t stationary? Justify your answer.

Problem 5: Autoregressive Models (Version C)

Consider a sequential process described by the following autoregressive model.

$$X_t = X_{t-1} + \varepsilon_t$$

The noise variables ε_t are independently $\mathcal{N}(0, 1)$ distributed and we condition the process on $X_0 = 0$.

0
1
2

a) Is this autoregressive process stationary? Justify your answer.

0
1
2

b) Now we define a second process Y_t as following

$$Y_t = X_t - X_{t-1}.$$

Is the process Y_t stationary? Justify your answer.

Problem 5: Autoregressive Models (Version D)

Consider a sequential process described by the following autoregressive model.

$$X_t = X_{t-1} + \varepsilon_t$$

The noise variables ε_t are independently $\mathcal{N}(0, 1)$ distributed and we condition the process on $X_0 = 0$.

a) Is this autoregressive process stationary? Justify your answer.

☐ 0
☐ 1
☐ 2

b) Now we define a second process Y_t as following

☐ 0
☐ 1
☐ 2

$$Y_t = X_t - X_{t-1}.$$

Is the process Y_t stationary? Justify your answer.

Problem 6: Hidden Markov Models (Version A)

Consider a hidden Markov model with 3 states 1, 2, 3. There are 3 possible observations a, b, c. The initial distribution π, transition probabilities $\boldsymbol{A}$ and emission probabilities $\boldsymbol{B}$ are

$$\pi = \begin{array}{c} 1 \\ 2 \\ 3 \end{array}\begin{pmatrix} 2/5 \\ 2/5 \\ 1/5 \end{pmatrix} \qquad \boldsymbol{A} = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 2/5 & 1/5 & 2/5 \\ 2 & 2/5 & 2/5 & 1/5 \\ 3 & 1/5 & 0 & 4/5 \end{array} \qquad \boldsymbol{B} = \begin{array}{c|ccc} & \mathrm{a} & \mathrm{b} & \mathrm{c} \\ \hline 1 & 3/5 & 1/5 & 1/5 \\ 2 & 1/5 & 3/5 & 1/5 \\ 3 & 2/5 & 0 & 3/5 \end{array},$$

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

0
1
2
3
4
5
6
7
8

We have observed the sequence $X_{1:3} = [\mathrm{bbc}]$. What is the most likely latent state Z_2 given these observations? Justify your answer.

$$\beta_t = A(B \odot \beta_{t+1})$$

$$\beta_2 = \begin{pmatrix} \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \\ \frac{1}{5} & 0 & \frac{4}{5} \end{pmatrix} \left(\begin{bmatrix} \frac{1}{5} \\ \frac{1}{5} \\ \frac{3}{5} \end{bmatrix} \odot 1 \right)$$

$$= \begin{pmatrix} \frac{9}{25} \\ \frac{7}{25} \\ \frac{13}{25} \end{pmatrix}$$

$$\alpha_1 = \pi \odot B$$

$$= \begin{pmatrix} \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{5} \end{pmatrix} \odot \begin{pmatrix} \frac{1}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{2}{25} \\ \frac{6}{25} \\ 0 \end{pmatrix}$$

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

$$= \begin{pmatrix} \frac{1}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix} \odot \begin{pmatrix} \frac{2}{5} & \frac{2}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{2}{5} & 0 \\ \frac{2}{5} & \frac{1}{5} & \frac{4}{5} \end{pmatrix} \begin{pmatrix} \frac{2}{25} \\ \frac{6}{25} \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix} \odot \begin{pmatrix} \frac{16}{125} \\ \frac{14}{125} \\ \dots \end{pmatrix} = \begin{pmatrix} \frac{16}{625} \\ \frac{42}{625} \\ 0 \end{pmatrix}$$

$$\alpha_2 \cdot \beta_2 = \begin{pmatrix} \frac{144}{15625} \\ \frac{294}{15625} \\ 0 \end{pmatrix}$$

$$Z_2 = 2$$

Problem 6: Hidden Markov Models (Version B)

Consider a hidden Markov model with 3 states 1, 2, 3. There are 3 possible observations a, b, c. The initial distribution $\boldsymbol{\pi}$, transition probabilities $\boldsymbol{A}$ and emission probabilities $\boldsymbol{B}$ are

$$\boldsymbol{\pi} = \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{pmatrix} 2/5 \\ 1/5 \\ 2/5 \end{pmatrix} \qquad \boldsymbol{A} = \begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 2/5 & 2/5 & 1/5 \\ 1/5 & 4/5 & 0 \\ 2/5 & 1/5 & 2/5 \end{pmatrix} \end{array} \qquad \boldsymbol{B} = \begin{array}{cc} & \begin{array}{ccc} \mathrm{a} & \mathrm{b} & \mathrm{c} \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 1/5 & 1/5 & 3/5 \\ 0 & 3/5 & 2/5 \\ 3/5 & 1/5 & 1/5 \end{pmatrix} \end{array},$$

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

We have observed the sequence $X_{1:3} = [\mathrm{aab}]$. What is the most likely latent state Z_2 given these observations? Justify your answer.

0 1 2 3 4 5 6 7 8

Problem 6: Hidden Markov Models (Version C)

Consider a hidden Markov model with 3 states 1, 2, 3. There are 3 possible observations a, b, c. The initial distribution $\boldsymbol{\pi}$, transition probabilities $\boldsymbol{A}$ and emission probabilities $\boldsymbol{B}$ are

$$\boldsymbol{\pi} = \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{pmatrix} 1/5 \\ 2/5 \\ 2/5 \end{pmatrix} \qquad \boldsymbol{A} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 4/5 & 0 & 1/5 \\ 1/5 & 2/5 & 2/5 \\ 2/5 & 1/5 & 2/5 \end{pmatrix} \end{array} \qquad \boldsymbol{B} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{a} & \mathrm{b} & \mathrm{c} \end{array} \\ \begin{pmatrix} 2/5 & 3/5 & 0 \\ 1/5 & 1/5 & 3/5 \\ 3/5 & 1/5 & 1/5 \end{pmatrix} \end{array},$$

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

0 1 2 3 4 5 6 7 8

We have observed the sequence $X_{1:3} = [\mathrm{ccb}]$. What is the most likely latent state Z_2 given these observations? Justify your answer.

Problem 6: Hidden Markov Models (Version D)

Consider a hidden Markov model with 3 states 1, 2, 3. There are 3 possible observations a, b, c. The initial distribution $\boldsymbol{\pi}$, transition probabilities $\boldsymbol{A}$ and emission probabilities $\boldsymbol{B}$ are

$$\boldsymbol{\pi} = \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{pmatrix} 2/5 \\ 2/5 \\ 1/5 \end{pmatrix} \qquad \boldsymbol{A} = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 2/5 & 1/5 & 2/5 \\ 2 & 2/5 & 2/5 & 1/5 \\ 3 & 1/5 & 0 & 4/5 \end{array} \qquad \boldsymbol{B} = \begin{array}{c|ccc} & \mathrm{a} & \mathrm{b} & \mathrm{c} \\ \hline 1 & 1/5 & 3/5 & 1/5 \\ 2 & 1/5 & 1/5 & 3/5 \\ 3 & 3/5 & 2/5 & 0 \end{array},$$

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

We have observed the sequence $X_{1:3} = [\mathrm{cca}]$. What is the most likely latent state Z_2 given these observations? Justify your answer.

0 1 2 3 4 5 6 7 8

Problem 7: Temporal point processes (Version A)

0
1
2
3
4
5
6

We simulate a homogeneous Poisson process with intensity μ on the interval $[0, T]$. What is the probability that we observe no events (i.e., the first event t_1 will hapen after T)? Justify your answer.

$$\lambda^*(t) = \mu$$

$$\# \int_0^T \lambda(t)dt = \mu t \Big|_0^T = \mu T$$

Only if $\mu = 0$

$$p(X=0) = \frac{e^{-\mu T}(\mu T)^0}{0!} = e^{-\mu T}$$

Problem 7: Temporal point processes (Version B)

We simulate a homogeneous Poisson process with intensity μ on the interval $[0, T]$. What is the probability that we observe no events (i.e., the first event t_1 will hapen after T)? Justify your answer.

0
1
2
3
4
5
6

Problem 7: Temporal point processes (Version C)

0 1 2 3 4 5 6

We simulate a homogeneous Poisson process with intensity μ on the interval [0, T]. What is the probability that we observe no events (i.e., the first event t_1 will hapen after T)? Justify your answer.

Problem 7: Temporal point processes (Version D)

We simulate a homogeneous Poisson process with intensity μ on the interval $[0, T]$. What is the probability that we observe no events (i.e., the first event t_1 will hapen after T)? Justify your answer.

0
1
2
3
4
5
6

Problem 8: Graphs - Clustering (Version A)

0
1
2
3
4
5
6

In this task, we consider a Stochastic Block Model with 2 communities that can generate *weighted* graphs. We assume that the edge weights are drawn from the Poisson distribution, that is

$$p(A_{ij} = k | z_i, z_j, \boldsymbol{\eta}) = \frac{\eta_{z_i z_j}^k}{k!} \exp(-\eta_{z_i z_j})$$

Suppose we observed the weighted adjacency matrix $\boldsymbol{A} \in \{0, 1, 2, ...\}^{N \times N}$ and the cluster indicators $\boldsymbol{z} \in \{1, 2\}^N$. Derive the maximim likelihood estimate of the parameters $\boldsymbol{\eta} = \begin{pmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{pmatrix}$.

Hint: It might be helpful to introduce shorthand notation

$$N_p = \sum_{i=1}^{N} \mathbb{1}(z_i = p) \qquad \textit{and} \qquad M_{pq} = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \mathbb{1}(z_i = p, z_j = q),$$

where $\mathbb{1}(\cdot)$ *is the indicator function.*

$$\log p = k \log \eta_{z_i z_j} - \log k! - \eta_{z_i z_j}$$

$$\sum_{ij} = \sum_{ij} k \log \eta_{z_i z_j} \underbrace{- \sum \log k!}_{C} - \sum \eta_{z_i z_j}$$

$$\Rightarrow \quad \sum_{pq} m_{pq} \log \eta_{pq} - \sum n_p n_q \cdot \eta_{pq}$$

$$\frac{\partial \cdots}{\eta_{pq}} = \frac{m_{pq}}{\eta_{pq}} - n_p n_q = 0 \qquad \eta_{pq} = \frac{m_{pq}}{n_p n_q}$$

Problem 8: Graphs - Clustering (Version B)

In this task, we consider a Stochastic Block Model with 2 communities that can generate *weighted* graphs. We assume that the edge weights are drawn from the Poisson distribution, that is

0 1 2 3 4 5 6

$$p(A_{ij} = k | z_i, z_j, \boldsymbol{\eta}) = \frac{\eta_{z_i z_j}^k}{k!} \exp(-\eta_{z_i z_j})$$

Suppose we observed the weighted adjacency matrix $\boldsymbol{A} \in \{0, 1, 2, ...\}^{N \times N}$ and the cluster indicators $\boldsymbol{z} \in \{1, 2\}^N$. Derive the maximim likelihood estimate of the parameters $\boldsymbol{\eta} = \begin{pmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{pmatrix}$.

Hint: It might be helpful to introduce shorthand notation

$$N_p = \sum_{i=1}^{N} \mathbb{1}(z_i = p) \qquad \textit{and} \qquad M_{pq} = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \mathbb{1}(z_i = p, z_j = q),$$

where $\mathbb{1}(\cdot)$ *is the indicator function.*

Problem 8: Graphs - Clustering (Version C)

0 1 2 3 4 5 6

In this task, we consider a Stochastic Block Model with 2 communities that can generate *weighted* graphs. We assume that the edge weights are drawn from the Poisson distribution, that is

$$p(A_{ij} = k | z_i, z_j, \boldsymbol{\eta}) = \frac{\eta_{z_i z_j}^k}{k!} \exp(-\eta_{z_i z_j})$$

Suppose we observed the weighted adjacency matrix $\boldsymbol{A} \in \{0, 1, 2, ...\}^{N \times N}$ and the cluster indicators $\boldsymbol{z} \in \{1, 2\}^N$. Derive the maximim likelihood estimate of the parameters $\boldsymbol{\eta} = \begin{pmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{pmatrix}$.

Hint: It might be helpful to introduce shorthand notation

$$N_p = \sum_{i=1}^{N} \mathbb{1}(z_i = p) \qquad \textit{and} \qquad M_{pq} = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \mathbb{1}(z_i = p, z_j = q),$$

where $\mathbb{1}(\cdot)$ *is the indicator function.*

Problem 8: Graphs - Clustering (Version D)

In this task, we consider a Stochastic Block Model with 2 communities that can generate *weighted* graphs. We assume that the edge weights are drawn from the Poisson distribution, that is

0 1 2 3 4 5 6

$$p(A_{ij} = k | z_i, z_j, \boldsymbol{\eta}) = \frac{\eta_{z_i z_j}^k}{k!} \exp(-\eta_{z_i z_j})$$

Suppose we observed the weighted adjacency matrix $\boldsymbol{A} \in \{0, 1, 2, ...\}^{N \times N}$ and the cluster indicators $\boldsymbol{z} \in \{1, 2\}^N$. Derive the maximim likelihood estimate of the parameters $\boldsymbol{\eta} = \begin{pmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{pmatrix}$.

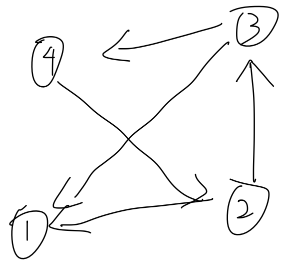
Hint: It might be helpful to introduce shorthand notation

$$N_p = \sum_{i=1}^{N} \mathbb{1}(z_i = p) \qquad \textit{and} \qquad M_{pq} = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} \mathbb{1}(z_i = p, z_j = q),$$

where $\mathbb{1}(\cdot)$ *is the indicator function.*

Problem 9: Graphs - Ranking (Version A)

We consider a directed graph $G = (V, E)$ with adjacency matrix



$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

where A_{ij} indicates if there exists an edge from node i to node j.

0 1 2 3 a) Find the stationary distribution $\pi(\infty)$ associated with the random walk on graph G. Justify your answer.

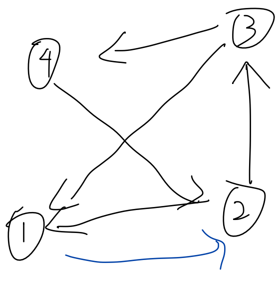
① without outgoing   normalized to 1

$\pi = [1, 0, 0, 0]$ $A\pi(\infty) = \pi(\infty)$

0 1 b) Explain why the Markov chain corresponding to the random walk on graph G is not irreducible.

① with out going dean end

0 1 2 3 4 5 6 c) We want to add a single directed edge (u, v) to the graph G and create a new graph $G' = (V, E \cup \{(u, v)\})$, such that the Markov chain corresponding to the random walk on graph G' becomes irreducible. What edge (u, v) should we add such that the PageRank score for node 1 in G' is also maximized? Also, compute the PageRank scores for all the nodes in G' after adding the chosen edge (u, v).



$r_1 = \frac{r_2}{2} + \frac{r_3}{2}$

$r_2 = r_1 + r_4$

$r_3 = \frac{r_2}{2}$

$r_4 = \frac{r_3}{2}$

$\sum r_i = 1$

$\frac{r_2}{2} + \frac{r_3}{2} + r_2 + r_3 + r_4 = 1$

$\frac{3}{2} r_2 + \frac{3}{2} r_3 + r_4 = 1$

$3r_3 + 3r_4 + r_4 = 1$

$6r_4 + 4r_4 = 1$

$r_4 = \frac{1}{10}$

Problem 9: Graphs - Ranking (Version B)

We consider a directed graph $G = (V, E)$ with adjacency matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

where A_{ij} indicates if there exists an edge from node i to node j.

a) Find the stationary distribution $\pi(\infty)$ associated with the random walk on graph G. Justify your answer.

0 1 2 3

b) Explain why the Markov chain corresponding to the random walk on graph G is not irreducible.

0 1

c) We want to add a single directed edge (u, v) to the graph G and create a new graph $G' = (V, E \cup \{(u, v)\})$, such that the Markov chain corresponding to the random walk on graph G' becomes irreducible. What edge (u, v) should we add such that the PageRank score for node 1 in G' is also maximized? Also, compute the PageRank scores for all the nodes in G' after adding the chosen edge (u, v).

0 1 2 3 4 5 6

Problem 9: Graphs - Ranking (Version C)

We consider a directed graph $G = (V, E)$ with adjacency matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

where A_{ij} indicates if there exists an edge from node i to node j.

0 1 2 3

a) Find the stationary distribution $\boldsymbol{\pi}(\infty)$ associated with the random walk on graph G. Justify your answer.

0 1

b) Explain why the Markov chain corresponding to the random walk on graph G is not irreducible.

0 1 2 3 4 5 6

c) We want to add a single directed edge (u, v) to the graph G and create a new graph $G' = (V, E \cup \{(u, v)\})$, such that the Markov chain corresponding to the random walk on graph G' becomes irreducible. What edge (u, v) should we add such that the PageRank score for node 1 in G' is also maximized? Also, compute the PageRank scores for all the nodes in G' after adding the chosen edge (u, v).

Problem 9: Graphs - Ranking (Version D)

We consider a directed graph $G = (V, E)$ with adjacency matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

where A_{ij} indicates if there exists an edge from node i to node j.

a) Find the stationary distribution $\pi(\infty)$ associated with the random walk on graph G. Justify your answer.

0
1
2
3

b) Explain why the Markov chain corresponding to the random walk on graph G is not irreducible.

0
1

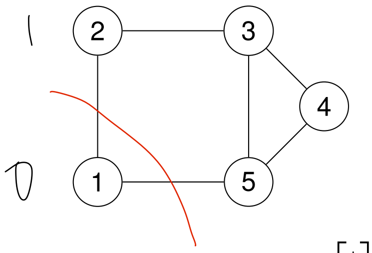
c) We want to add a single directed edge (u, v) to the graph G and create a new graph $G' = (V, E \cup \{(u, v)\})$, such that the Markov chain corresponding to the random walk on graph G' becomes irreducible. What edge (u, v) should we add such that the PageRank score for node 1 in G' is also maximized? Also, compute the PageRank scores for all the nodes in G' after adding the chosen edge (u, v).

0
1
2
3
4
5
6

Problem 10: Graphs - Semi-Supervised Learning (Version A)

We consider the graph G with the following adjacency matrix $\boldsymbol{A}$.

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$



We assume that the vector of labels is $\boldsymbol{y} = \begin{bmatrix} \hat{\boldsymbol{y}}_S \\ \boldsymbol{y}_U \end{bmatrix}$ where $\hat{\boldsymbol{y}}_S = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ are known labels and $\boldsymbol{y}_U = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$ are unknown labels.

In other words, $S = \{1, 2\}$ is the set of labeled nodes and $U = \{3, 4, 5\}$ is the set of unlabeled nodes.

0 1 2 3

a) Find the optimal cost and all the optimal solutions to the following optimization problem:

$$\begin{aligned} \boldsymbol{y}^* &= \arg\min_{\boldsymbol{y} \in \{0,1\}^5} \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} \\ &\text{subject to } y_i = \hat{y}_{S,i} \text{ for } i \in \{1, 2\} \end{aligned} \tag{10.1}$$

where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ denotes the Laplacian of the graph G, and $\boldsymbol{D}$ is the degree matrix. Justify your answer.

$$D = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} \qquad L = \begin{pmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & -1 & 3 \end{pmatrix}$$

= minimum cut

2 3 4 5 ⇒ 1

or 1 3 4 5 ⇒

0 1 2 3 4 5

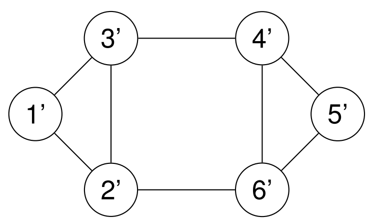
b) We consider a relaxation of the optimization problem from Equation (10.1) (note that we now optimize over $\mathbb{R}^6$):

$$\begin{aligned} \boldsymbol{y}^* &= \arg\min_{\boldsymbol{y} \in \mathbb{R}^5} \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} \\ &\text{subject to } y_i = \hat{y}_{S,i} \text{ for } i \in \{1, 2\} \end{aligned} \tag{10.2}$$

Suppose that the solution $\boldsymbol{y}^*$ of this optimization problem is given.

We now consider a modified graph G' with adjacency matrix

$$\boldsymbol{A}' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$



We denote the labels of G' as $\boldsymbol{y}' = \begin{bmatrix} \hat{\boldsymbol{y}}'_S \\ \boldsymbol{y}'_U \end{bmatrix}$, where $\boldsymbol{y}'_S = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ are known labels and $\boldsymbol{y}'_U = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$ are unknown labels.

Your task is to explain how the solution $\boldsymbol{y}^*$ of the optimization problem in Equation 10.2 can be used to solve the following optimization problem:

$$\begin{aligned} &\arg\min_{\boldsymbol{y}' \in \mathbb{R}^6} \boldsymbol{y}'^T \boldsymbol{L}' \boldsymbol{y}' \\ &\text{subject to } y'_i = \hat{y}'_{S,i} \text{ for } i \in \{1, 2, 3\} \end{aligned} \tag{10.3}$$

where $\boldsymbol{L}' = \boldsymbol{A}' - \boldsymbol{D}'$ is the Laplacian of G'. Justify your answer.

$$y_u = -L_{uu}^{-1} L_{us} \hat{y}_s$$

$$L_{uu} = L'_{uu}$$

$$\therefore \; L'_{us} = [0 \quad L_{us}]$$

$$y'^*_u = -L_{uu}^{-1} \underline{[0 \quad L_{us}] \hat{y}'_s} \approx \begin{pmatrix} 1 \\ b \end{pmatrix}$$

$$-L_{uu}^{-1} L_{us} \begin{pmatrix} 1 \\ b \end{pmatrix}$$

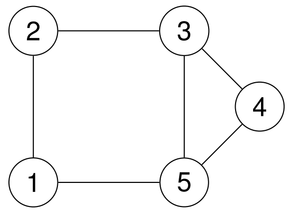
$$= y^*_u$$

$$y'^* = \begin{bmatrix} 1 \\ y^* \end{bmatrix}$$

Problem 10: Graphs - Semi-Supervised Learning (Version B)

We consider the graph G with the following adjacency matrix $\boldsymbol{A}$.

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$



We assume that the vector of labels is $\boldsymbol{y} = \begin{bmatrix} \hat{\boldsymbol{y}}_S \\ \boldsymbol{y}_U \end{bmatrix}$ where $\hat{\boldsymbol{y}}_S = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ are known labels and $\boldsymbol{y}_U = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$ are unknown labels.
In other words, $S = \{1, 2\}$ is the set of labeled nodes and $U = \{3, 4, 5\}$ is the set of unlabeled nodes.

a) Find the optimal cost and <u>all</u> the optimal solutions to the following optimization problem:

$$\begin{aligned} \boldsymbol{y}^* &= \arg\min_{\boldsymbol{y} \in \{0,1\}^5} \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} \\ &\text{subject to } y_i = \hat{y}_{S,i} \text{ for } i \in \{1, 2\} \end{aligned} \tag{10.1}$$

where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ denotes the Laplacian of the graph G, and $\boldsymbol{D}$ is the degree matrix. Justify your answer.

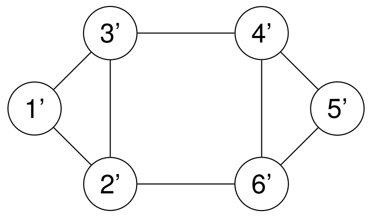
b) We consider a relaxation of the optimization problem from Equation (10.1) (note that we now optimize over $\mathbb{R}^6$):

$$\begin{aligned} \boldsymbol{y}^* &= \arg\min_{\boldsymbol{y} \in \mathbb{R}^5} \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} \\ &\text{subject to } y_i = \hat{y}_{S,i} \text{ for } i \in \{1, 2\} \end{aligned} \tag{10.2}$$

Suppose that the solution $\boldsymbol{y}^*$ of this optimization problem is given.

We now consider a modified graph G' with adjacency matrix

$$\boldsymbol{A}' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$



We denote the labels of G' as $\boldsymbol{y}' = \begin{bmatrix} \hat{\boldsymbol{y}}'_S \\ \boldsymbol{y}'_U \end{bmatrix}$, where $\boldsymbol{y}'_S = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ are known labels and $\boldsymbol{y}'_U = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$ are unknown labels.

Your task is to explain how the solution $\boldsymbol{y}^*$ of the optimization problem in Equation 10.2 can be used to solve the following optimization problem:

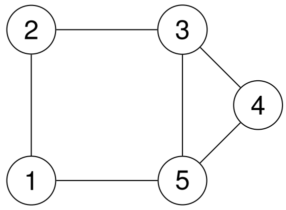
$$\begin{aligned} &\arg\min_{\boldsymbol{y}' \in \mathbb{R}^6} \boldsymbol{y}'^T \boldsymbol{L}' \boldsymbol{y}' \\ &\text{subject to } y'_i = \hat{y}'_{S,i} \text{ for } i \in \{1, 2, 3\} \end{aligned} \tag{10.3}$$

where $\boldsymbol{L}' = \boldsymbol{A}' - \boldsymbol{D}'$ is the Laplacian of G'. Justify your answer.

Problem 10: Graphs - Semi-Supervised Learning (Version C)

We consider the graph G with the following adjacency matrix $\boldsymbol{A}$.

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$



We assume that the vector of labels is $\boldsymbol{y} = \begin{bmatrix} \hat{\boldsymbol{y}}_S \\ \boldsymbol{y}_U \end{bmatrix}$ where $\hat{\boldsymbol{y}}_S = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ are known labels and $\boldsymbol{y}_U = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$ are unknown labels.
In other words, $S = \{1, 2\}$ is the set of labeled nodes and $U = \{3, 4, 5\}$ is the set of unlabeled nodes.

a) Find the optimal cost and <u>all</u> the optimal solutions to the following optimization problem:

$$\begin{aligned} \boldsymbol{y}^* = \arg\min_{\boldsymbol{y} \in \{0,1\}^5} & \ \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} \\ \text{subject to } & \ y_i = \hat{y}_{S,i} \text{ for } i \in \{1, 2\} \end{aligned} \tag{10.1}$$

where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ denotes the Laplacian of the graph G, and $\boldsymbol{D}$ is the degree matrix. Justify your answer.

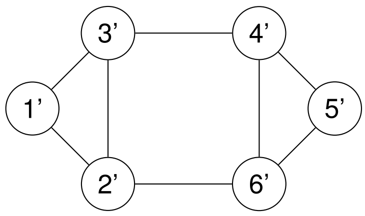
b) We consider a relaxation of the optimization problem from Equation (10.1) (note that we now optimize over $\mathbb{R}^6$):

$$\begin{aligned} \boldsymbol{y}^* = \arg\min_{\boldsymbol{y} \in \mathbb{R}^5} & \ \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} \\ \text{subject to } & \ y_i = \hat{y}_{S,i} \text{ for } i \in \{1, 2\} \end{aligned} \tag{10.2}$$

Suppose that the solution $\boldsymbol{y}^*$ of this optimization problem is given.

We now consider a modified graph G' with adjacency matrix

$$\boldsymbol{A}' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$



We denote the labels of G' as $\boldsymbol{y}' = \begin{bmatrix} \hat{\boldsymbol{y}}'_S \\ \boldsymbol{y}'_U \end{bmatrix}$, where $\boldsymbol{y}'_S = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ are known labels and $\boldsymbol{y}'_U = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$ are unknown labels.

Your task is to explain how the solution $\boldsymbol{y}^*$ of the optimization problem in Equation 10.2 can be used to solve the following optimization problem:

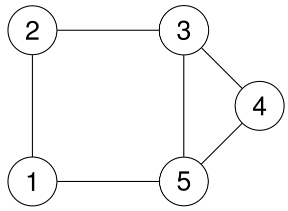
$$\begin{aligned} \arg\min_{\boldsymbol{y}' \in \mathbb{R}^6} & \ \boldsymbol{y}'^T \boldsymbol{L}' \boldsymbol{y}' \\ \text{subject to } & \ y'_i = \hat{y}'_{S,i} \text{ for } i \in \{1, 2, 3\} \end{aligned} \tag{10.3}$$

where $\boldsymbol{L}' = \boldsymbol{A}' - \boldsymbol{D}'$ is the Laplacian of G'. Justify your answer.

Problem 10: Graphs - Semi-Supervised Learning (Version D)

We consider the graph G with the following adjacency matrix $\boldsymbol{A}$.

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$



We assume that the vector of labels is $\boldsymbol{y} = \begin{bmatrix} \hat{\boldsymbol{y}}_S \\ \boldsymbol{y}_U \end{bmatrix}$ where $\hat{\boldsymbol{y}}_S = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ are known labels and $\boldsymbol{y}_U = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$ are unknown labels.
In other words, $S = \{1, 2\}$ is the set of labeled nodes and $U = \{3, 4, 5\}$ is the set of unlabeled nodes.

a) Find the optimal cost and <u>all</u> the optimal solutions to the following optimization problem:

$$\begin{aligned} \boldsymbol{y}^* &= \arg\min_{\boldsymbol{y} \in \{0,1\}^5} \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} \\ &\text{subject to } y_i = \hat{y}_{S,i} \text{ for } i \in \{1, 2\} \end{aligned} \tag{10.1}$$

where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ denotes the Laplacian of the graph G, and $\boldsymbol{D}$ is the degree matrix. Justify your answer.

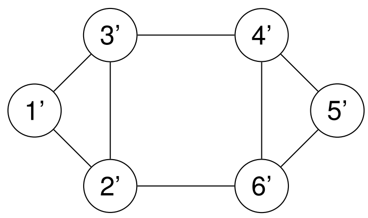
b) We consider a relaxation of the optimization problem from Equation (10.1) (note that we now optimize over $\mathbb{R}^6$):

$$\begin{aligned} \boldsymbol{y}^* &= \arg\min_{\boldsymbol{y} \in \mathbb{R}^5} \boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y} \\ &\text{subject to } y_i = \hat{y}_{S,i} \text{ for } i \in \{1, 2\} \end{aligned} \tag{10.2}$$

Suppose that the solution $\boldsymbol{y}^*$ of this optimization problem is given.

We now consider a modified graph G' with adjacency matrix

$$\boldsymbol{A}' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$



We denote the labels of G' as $\boldsymbol{y}' = \begin{bmatrix} \hat{\boldsymbol{y}}'_S \\ \boldsymbol{y}'_U \end{bmatrix}$, where $\boldsymbol{y}'_S = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ are known labels and $\boldsymbol{y}'_U = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$ are unknown labels.

Your task is to explain how the solution $\boldsymbol{y}^*$ of the optimization problem in Equation 10.2 can be used to solve the following optimization problem:

$$\begin{aligned} &\arg\min_{\boldsymbol{y}' \in \mathbb{R}^6} \boldsymbol{y}'^T \boldsymbol{L}' \boldsymbol{y}' \\ &\text{subject to } y'_i = \hat{y}'_{S,i} \text{ for } i \in \{1, 2, 3\} \end{aligned} \tag{10.3}$$

where $\boldsymbol{L}' = \boldsymbol{A}' - \boldsymbol{D}'$ is the Laplacian of G'. Justify your answer.