Algorithmic Foundations of Data Science

Sample Exam, solve whenever you want

Name: ____________________

Student ID: ____________________

Study Program: ____________________

Remarks

- Write your name and student ID on **every** sheet of paper.
- Write your solution in the space provided on the problem sheets. You may use the back of the pages if you need extra space. Please mark it below the problem if you do so.
- Do **not** use your **own paper**.
- You may answer in either English or German. Please do not mix languages within the answer to a problem.
- Only use **black** or **blue** document-proof pens. Do **not** use pencils.
- Clearly mark your solutions and results as such. If you provide **multiple** solutions to a question, the **worst** of those counts.
- You have **120 minutes** to work on the exam.
- With **60 points** you have passed this exam.

I hereby declare that I have read the above guidelines and that I am healthy enough to take the exam.

____

(Signature)

Do not write below this line.

	1	**2**	**3**	**4**	**5**	**6**
Points	/ 20	/ 20	/ 21	/ 19	/ 19	/ 21
Sign.						

$\sum$	/ 120

Problem 1 (General Questions) 4+4+4+4+4 = 20 points

a) Briefly explain the concepts of batch and online learning in the context of supervised learning. What is the difference?

b) What is the property of the hypothesis returned by an Empirical Risk Minimization algorithm?

c) Where is the volume of high-dimensional balls concentrated? Name the intuitive meaning of both bounds we considered in the lecture.

d) Name two applications of the Multiplicative Weight Update Algorithm discussed in the lecture.

e) Sketch the Page Rank algorithm.

Problem 2 (Markov Chains) 5+5+5+5 = 20 points

a) Let $\mathcal{Q}$ be a Markov chain and let G_Q denote its graph. Give the (formal) definitions of (i) *connectedness* of $\mathcal{Q}$, (ii) *aperiodicity* of $\mathcal{Q}$, and (iii) *ergodicity* of $\mathcal{Q}$. Additionally, give an example of an ergodic Markov chain with at least three states.

b) Give an example of a *connected* Markov chain $\mathcal{Q}$ with transition matrix Q such that there exists an initial distribution $\boldsymbol{p}_0$ for which the sequence $\left(\boldsymbol{p}_0 Q^i\right)_{i\in\mathbb{N}}$ does *not* converge to the the stationary distribution of $\mathcal{Q}$. For this, determine the stationary distribution $\boldsymbol{\pi}$ of your chain $\mathcal{Q}$ and briefly argue why $\lim_{i\to\infty} \boldsymbol{p}_0 Q^i \neq \boldsymbol{\pi}$.

c) You are observing traffic on the A4 motorway at the border from Germany to the Netherlands. There are only cars and trucks on the road. You find that on average

- for five in six cars crossing the border, the next vehicle crossing the border is again a car;
- for every second truck crossing the border, the next vehicle crossing the border is again a truck.

Model this situation with a two state Markov chain (state 1 $\widehat{=}$ car, state 2 $\widehat{=}$ truck). Compute the stationary distribution of your Markov chain to find the total percentage of cars and trucks on the road.

d) Let $Q = (q_{ij}) \in \mathbb{R}^{n \times n}$ be the transition matrix of a connected Markov chain $\mathcal{Q}$. We construct a new matrix $Q' = (q'_{ij}) \in \mathbb{R}^{n \times n}$ with

$$q'_{ij} := \begin{cases} \frac{1}{2} q_{ij} & \text{if } i \neq j \\ \frac{1}{2}(q_{ij} + 1) & \text{if } i = j. \end{cases}$$

Prove that Q' is the transition matrix of an *ergodic* Markov chain $\mathcal{Q}'$. Prove that $\mathcal{Q}'$ has the same stationary distribution as $\mathcal{Q}$.

Problem 3 (Streaming) 4+2+3+6+6 = 21 points

a) Let $p \geq 0$. Define the pth frequency moment $F_p(\boldsymbol{a})$ of a data stream $\boldsymbol{a} = a_1, \ldots, a_n$ of elements from a universe U. Which special restriction is needed for $p = 0$?

b) What is the intuitive meaning of $F_0(\boldsymbol{a})$ and $F_1(\boldsymbol{a})$?

c) Compute $F_0(\boldsymbol{a})$, $F_1(\boldsymbol{a})$, and $F_2(\boldsymbol{a})$ for the stream $\boldsymbol{a} = 2, 3, 1, 6, 1, 2, 6, 3, 2, 7, 1$ from $U = [8]$.

d) Describe the Flajolet-Martin Algorithm for estimating the number of distinct elements of the stream. Which properties of the family of hash functions used by the Algorithm are assumed?

e) What is the result of applying the Flajolet-Martin Algorithm to the stream in (c) if the randomly chosen hash function $h\colon U \to [8]$ is given below. Give the result as well as some intermediate steps showing how the estimator is computed from $\boldsymbol{a}$

a	1	2	3	4	5	6	7	8
$h(a)$	6	5	4	8	7	2	1	3

Problem 4 (Power Iteration Algorithm) 6+7+6 = 19 points

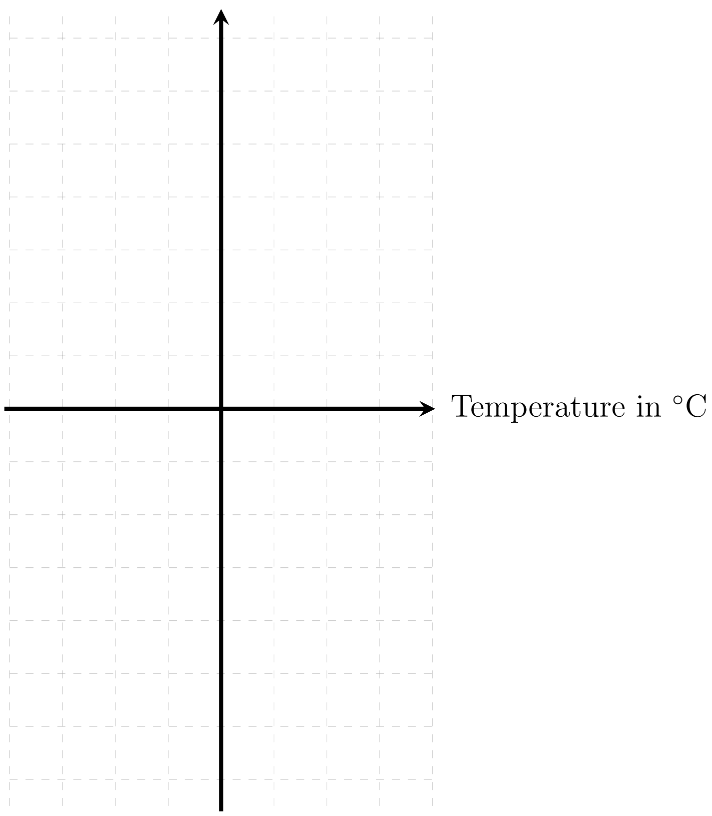
The local ice cream shop wants to find out how their sells depend on the weather. Towards this goal they measure the temperature on four days and count how many scoops of ice cream they sell. They find that the mean temperature was 30°C and the mean of sold scoops was 700 per day.

The following table notes the deviation from the mean on each day.

	Temperature (in °C)	Scoops of ice cream (in increments of 10)
day 1	-2	-3
day 2	3	7
day 3	2	3
day 4	-3	-7

Let A be the 4×2 matrix representing these data-points.

a) Plot the 4 data-points above on the coordinate system drawn below. Using this plot, graphically *estimate* the first principal component of the above data-set. State this estimate as a linear equation of the form $ax + by = c$.



b) Recall that the co-variance matrix C of the data-set A is defined as $C = A^{\mathsf{T}}A$. Using $(1 \quad 1)^{\mathsf{T}}$ as the starting vector, apply 2 rounds of the Power Iteration algorithm to estimate the eigenvector of C corresponding to the largest eigenvalue. To make the computations easier do the following:

- Use the L^{∞}-norm (i.e., the absolute value of the largest entry in the vector) for normalization.
- Round the vectors to one digit after the decimal point, at *every* step of Power Iteration.

c) Let v be the eigenvector estimated by you in Part b). Let $\langle v \rangle$ denote the line in the x-y plane which passes through $(0, 0)^\mathsf{T}$ and v. Compare $\langle v \rangle$ with the estimate obtained in Part a) for the first principal component. Are they similar? If yes, briefly explain why it is reasonable to expect this. If no, give a reason why they are far from each other in this example.

Problem 5 (Map-Reduce) 3+10+6 = 19 points

a) Name and describe three cost measures that are used in the analysis of Map-Reduce algorithms.

b) We consider meteorological data, given in key-value pairs of the shape $(c, (s, t, d))$ where

- c is a country,
- s is a weather station,
- t is a temperature measurement (in °C) and
- d is the date of the recording.

Specify *single-round* Map-Reduce algorithms for the following problems in pseudocode.

(i) *Average temperature per country.* Output all key-value pairs (c, t) where c is a country and t is the average temperature in country c (taken over all measurements ever recorded in country c).

(ii) *Stations with extreme temperature differences.* Output all key-value pairs (c, s) where c is a country and s is a station in country c with the property that the difference between the lowest and the highest temperature ever recorded at station s is at least $30\,°\mathrm{C}$.

Additional notes:

- Use the "on input ..., [do some computation,] emit ..." format for specifying your pseudocode.
- Make sure your algorithm can benefit from parallelisation.

c) The following shows a Map-Reduce algorithm for computing the difference of two relations $\mathcal{R}$ and $\mathcal{S}$ in Relational Algebra.

MAP: On input (R, t), emit $(Q, (R, t))$.
On input (S, t), emit $(Q, (S, t))$.

REDUCE: On input (Q, values),
emit (Q, t) if $(R, t) \in \mathit{values}$ and $(S, t) \notin \mathit{values}$.

Although this algorithm is technically correct, it is a very bad example of a Map-Reduce algorithm. Why?

Propose a new algorithm for computing the difference. Explain why your algorithm is better by comparing it to the algorithm above.

Problem 6 (Linear Separators) 5+6+10 = 21 points

In this problem we study linear separators and Boolean functions $f\colon \{0,1\}^n \to \{-1,1\}$. Recall that we say f can be *represented by a linear separator* if there is a *weight* vector $\boldsymbol{w} \in \mathbb{R}^n$ and a *bias* $b \in \mathbb{R}$ such that for all $\boldsymbol{x} \in \{0,1\}^n$,

$$f(\boldsymbol{x}) = \operatorname{sgn}\big(\langle \boldsymbol{w} \rangle \boldsymbol{x} - b\big).$$

where $\operatorname{sgn}(0) = 0$, $\operatorname{sgn}(x) = 1$ if $x > 0$, and $\operatorname{sgn}(x) = -1$ if $x < 0$. In this case, we say $(\boldsymbol{w}, b)$ *represents* f.

a) Consider the function $f\colon \{0,1\}^2 \to \{-1,1\}$ that is defined as follows:

$\boldsymbol{x}$	$\binom{0}{0}$	$\binom{0}{1}$	$\binom{1}{0}$	$\binom{1}{1}$
$f(\boldsymbol{x})$	1	-1	1	-1

Find a pair $(\boldsymbol{w}, b)$ representing f.

b) We consider a generalisation of the function in part a): For all $n \in \mathbb{N}_{>0}$ and every $I \subseteq \{1, \dots, n\}$, we let $f_{n,I} \colon \{0,1\}^n \to \{-1, 1\}$ with

$$f_{n,I}(\boldsymbol{x}) \coloneqq \begin{cases} -1 & \text{if } x_i = 1 \text{ for all } i \in I \text{ and} \\ 1 & \text{otherwise} \end{cases}$$

for all $\boldsymbol{x} = (x_1, \dots, x_n)^\intercal \in \{0,1\}^n$.

Depending on n and I, find a pair $(\boldsymbol{w}_{n,I}, b_{n,I})$ representing $f_{n,I}$. Justify your solution by discussing the value of $\langle \boldsymbol{w}_{n,I} \rangle \boldsymbol{x} - b_{n,I}$ for $\boldsymbol{x} \in \{0,1\}^n$.

c) Let $n \in \mathbb{N}_{>0}$ and let $f_n \colon \{0,1\}^n \to \{-1,1\}$ be the Boolean function defined by

$$f_n(\boldsymbol{x}) \coloneqq \begin{cases} -1 & \text{if } 2x_1 + \sum_{i=2}^{n} 3x_i \equiv 0 \pmod 4 \\ 1 & \text{otherwise.} \end{cases}$$

(i) Find pairs $(\boldsymbol{w}, b)$ and $(\boldsymbol{w}', b')$ representing f_1 and f_2, respectively.

(ii) Show that for $n \geq 3$, f_n can not be represented by a linear separator.

Hint:. First show that f_3 has no linear separator and then generalise the proof to $n > 3$.

Empty Page for Notes