ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Foundations of Data Science
Fall 2022

Assignment date: Thursday, November 17th, 2022, 17:15
Due date: Thursday, November 17th, 2022, 19:00

Midterm Exam – INF1

This exam is open book. No electronic devices of any kind are allowed. There are five problems. Good luck!

Only answers given on this handout count.

Name: _____

Problem 1	/ 2
Problem 2	/ 5
Problem 3	/ 10
Problem 4	/ 10
Problem 5	/ 8
Total	/35

Problem 1 (Subgaussian RV)**.** [2pts; 0.5pts per question] Let X and Y be two independent subgaussian random variables. Which of the following are always subgaussian?

a) $X+Y$ b) X^2 c) XY d) $\begin{cases} 0, & \text{if } |X| \leq 1 \\ \frac{1}{|X|}, & \text{otherwise} \end{cases}$

(space for problem 1)

Problem 2 (Property Testing: Variance)**.** [5pts] A colleague claims to have implemented an algorithm which outputs i.i.d. samples distributed according to a discrete distribution P that has unit variance. Your task is to design a statistic to test whether this is indeed true.

Let Δ_k be the set of probability distributions on the alphabet $\mathcal{X} = \{1, \cdots, k\}$. Assume that $P \in \mathcal{P} \cup \mathcal{Q}$ with $\mathcal{P} := \{P \in \Delta_k : P \text{ has variance } 1\}$ and $\mathcal{Q} := \{P \in \Delta_k : P \text{ has variance } \in [0, 1-\epsilon] \cup [1+\epsilon, \infty)\}$, where $0 < \epsilon < 1$. You are given n samples $\{X_i\}_{i=1}^n$, where the X_i are independent copies sampled according to P.

Remark: For the following three questions we do not ask you to write down a proof (or explicit calculation) that your proposed solution works.

a) [1pt] We say that an estimator $e : S^n \mapsto \Pi$ on a sample S^n of length n (ϵ, δ)-learns a parameter $p \in \Pi$ if for any $(\epsilon, \delta) \in (0,1)^2$, given sufficiently many samples n, we have that $\mathbb{P}(\{|e(S^n) - p| > \epsilon\}) < \delta$. Give a brief explanation (one sentence, no calculations) why the empirical estimator of the second moment $\hat{\mu}_{X^2} := \frac{1}{n}\sum_{i=1}^n X_i^2$ can (ϵ, δ)-learn the second moment in our setting.

b) [2pts] First, assume that a genie tells you that X has zero mean. Design a simple test statistic and give a threshold in order to check for the above mentioned unit variance property.

Hint: Use the claim in a).

c) [2pts] Now consider the more general case where X can have arbitrary mean. Again, design a simple test statistic and give a threshold.

Hint: You can assume that $\hat{\mu}_X^2$ (ϵ, δ)-learns $\mathbb{E}[X]^2$.

(space for problem 2)

(space for problem 2)

(space for problem 2)

Problem 3 (Perfect Secrecy)**.** [10pts]

Alice is trying to communicate a message $M \in \mathcal{M}$ securely to Bob with the help of a shared secret key K. Alice *encrypts* her message into what is called a ciphertext C by computing $C = E_K(M)$. Here, $E_K(M)$ is a known function given the key K that takes the message M as input and outputs C (which in general is in a different alphabet that the message M).

Bob receives the message. He then *decrypts* the ciphertext by computing $M = D_K(C)$. Just as E_K, D_K is a known function for a fixed key K.

A cryptosystem is said to be *perfectly secure* if M and C are independent.

(a) [2pts] What are the values of $H(C|M,K), H(M|C,K)$ and $I(M;C)$ in a perfectly secure cryptosystem?

(b) [3pts] Show that $H(M|C) \leq H(K|C)$ for a perfectly secure cryptosystem.

(c) [3pts] Show that $H(K) \geq \log_2(|\mathcal{M}|))$ for a perfectly secure cryptosystem. *Hint:* First establish an inequality between $H(K)$ and $H(M)$.

(d) [2pts] During Roman times, the Caeser shift cipher was used to transmit messages securely. Every letter in the message was replaced by another letter some fixed number of positions away and the shift was decided by the key. For example, if the key was 3, the letter A would be replaced by D, B by E and so on. To describe this scheme mathematically map each letter to a number $A \to 0, B \to 1, \ldots, Z \to 25$. Then,

$$E_K(M) = (M + K) \bmod 26, \quad \text{and} \tag{1}$$
$$D_K(C) = (C_K) \bmod 26. \tag{2}$$

Show that the Caeser shift cipher is perfectly secure if the key K is distributed uniformly over $\mathcal{K} = \{0, 1, \ldots, 25\}$ and we are transmitting only a single letter.

(space for problem 3)

(space for problem 3)

(space for problem 3)

Problem 4 (Minimum-Norm Solutions)**.** [10pts] Let us consider an *underdetermined* system of linear equations $A\mathbf{x} = \mathbf{b}$, where $A \in \mathbb{R}^{m \times n}$ is a "fat" matrix ($m < n$) of rank m and where $\mathbf{b}$ is chosen such that a solution exists. In this case, there exist infinitely many solutions. Further, we denote the compact SVD of A by $A = U_{A,\text{red}}\Sigma_{A,\text{red}}V_{A,\text{red}}^T$ where $U_{A,\text{red}} \in \mathbb{R}^{m \times m}, \Sigma_A \in \mathbb{R}^{m \times m}, V_A \in \mathbb{R}^{n \times m}$.

a) [2pts] We first construct a "completed" system of equations

$$\tilde{A}\mathbf{x} = \tilde{\mathbf{b}} \tag{3}$$

with $\tilde{A} := \begin{bmatrix} A \\ B \end{bmatrix} \in \mathbb{R}^{n \times n}$ by appending $n - m$ rows to A. How do we have to choose the rows of B in order to guarantee that (3) has exactly one solution? Which dimensions do the components of the reduced SVD of $B = U_{B,\text{red}}\Sigma_{B,\text{red}}V_{B,\text{red}}^T$ have?

b) [3pts] How do the columns of $V_{B,\text{red}}$ relate to those of $V_{A,\text{red}}$? Restate the SVD of $\tilde{A}$ in terms of $U_{A,\text{red}}, \Sigma_{A,\text{red}}, V_{A,\text{red}}$, $U_{B,\text{red}}, \Sigma_{B,\text{red}}, V_{B,\text{red}}$ and then write down (3) in terms of this expression, also making explicit how $\tilde{b}$ relates to b and B.

c) [2pts] State a simple way to find *some* solution to the original problem through solving (3) for x.

d) [3pts] Finally, prove that the one solution $\mathbf{x}$ that has the minimum 2-norm can be expressed as

$$\mathbf{x}_{MN} = V_{A,red}\Sigma_{A,red}^{-1}U_{A,red}^T\mathbf{b}. \tag{4}$$

(space for problem 4)

(space for problem 4)

(space for problem 4)

Problem 5. *(Bandits)*[8pts] You commute every single day (weekends included) by bike between EPFL and Lausanne center (we only consider this one direction to keep things simple). There are many different paths you can take: along the lake, bike roads, through the center etc. Let the number of possible paths be K. Further, let's assume that you have two different bicycles. A racing bike and a mountain bike. You are trying to figure out the 'best' combination of route and bike, let's say according to the time it takes you to commute.

Every day you can measure the time. This time of course depends on the route you take and the bike you pick. But lets say that it also depends on whether it rains or not (to keep things simple assume that it rains roughly half the time) and what day of the week it is. In addition many small random and unknown factors influence the time you measure.

a) [2pts] Formulate this as a bandit problem.

b) [2pts] What algorithm do you choose and what regret do you expect to see.

Consider next the following variation. Everything is as before but there are two changes. First, you have an archrival who knows your policy and every day may place various obstacles on some of the paths. Assume that this is the dominant factor that determines the commute time and that the obstacles may increase the time it takes to commute up to a factor 2. Secondly, each day you have only one of the two bikes available to you (perhaps it is a publi-bike or you share).

c) [2pts] Formulate this as a bandit problem.

d) [2pts] What algorithm do you choose and what regret do you expect to see.

Note: There are several reasonable answers and many possible considerations. Hence make sure to justify why you make the choices you make.

(space for problem 5)

(space for problem 5)

(space for problem 5)