ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Foundations of Data Science — Assignment date: Friday, January 29th, 2021, 8:15
Fall 2020 — Due date: Friday, January 29th, 2021, 11:15

Final Exam – CM1121

This exam is open book. No electronic devices of any kind are allowed. There are six problems. We do not presume that you will finish all of them. Choose the ones you find easiest and collect as many points as possible. Good luck!

Name: ______________________________

Problem 1	/ 10
Problem 2	/ 15
Problem 3	/ 15
Problem 4	/ 10
Problem 5	/ 10
Problem 6	/ 10
Total	/70

Problem 1. *(These Bandits – 10 pts)*

Consider an adversarial bandit setting with K bandits, where the rewards are arbitrary numbers $x_{t,k} \in [0,1]$ (t stands for the time index and runs from 1 to n and k is the index of the bandit, which goes from 1 to K). You are the adversary, in charge of designing the rewards. You know that the policy that is used is the exp3 algorithm.

Your task is to fill in the numbers. You are given the constraint that the "average" value of all rewards must be $\frac{1}{2}$, where the "average" means the sum over all $n \times K$ entries divided by $n \times K$.

Your aim is to make the expected reward (not regret) of the player as small as possible.

(i) [5pts] In general, what is the expected reward the player gets in this adversarial setting when using the exp3 algorithm? State the reward normalized by the time n. We are only interested in the first order term, i.e., the constant, and not higher order terms that vanish with n.

(ii) [5pts] Explain how you fill in the numbers to minimize the expected reward and compute this reward. As before, our interest is in the first order term.

(space for problem 1)

(space for problem 1)

Problem 2. *(Estimating Entropy – 15pts)*

You are given n iid samples of a Bernoulli random variable with parameter μ. The parameter is known to be in the range $[\kappa, 1-\kappa]$, where $0 < \kappa \leq \frac{1}{2}$. Let the samples be denoted by $S = \{X_1, X_2, \cdots, X_n\}$, $X_i \in \{0, 1\}$, $i = 1, \cdots n$.

Your task is to estimate the entropy of the underlying distribution accurately. Let h denote the true entropy of the distribution and $\hat{h} = \hat{h}(S)$ be your estimate.

(i) [5pts] Design a scheme to accurately estimate h. Give an explicit epression for $\hat{h}$ as a function of the samples S.

(ii) [5pts] Since the samples S are random your estimate $\hat{h}(S)$ is a random variable. Let $\delta, \epsilon > 0$. Derive a bound of the form

$$\mathbb{P}\{|\hat{h}(S) - h| \geq \epsilon\} \leq \delta.$$

(iii) [5pts] In the expression of (ii) assume that you set δ to some fixed constant. How does the gap ϵ behave as a function of n?

Hint: Simple does it.

(space for problem 2)

(space for problem 2)

Problem 3. *(Compression: Fibonacci Coding – 15pts)*

Consider the following binary encoding of a positive integer n: $\mathcal{C}_F(n) = I_1 \dots I_r 1$, where $n = \sum_{i=1}^{r} I_i F_{i+1}$ and F_i is i-th Fibonacci number, $F_0 = 0$, $F_1 = 1$, $F_2 = F_0 + F_1 = 1$, $\cdots$, $F_i = F_{i-1} + F_{i-2}$, $i \geq 2$, and $I_i \in \{0, 1\}$. E.g., 1011 denotes the integer $1 \times 1 + 0 \times 2 + 1 \times 3 = 4$.

For every positive integer n such a representation exists. In order to make it unique, given an integer, find the largest Fibonacci number that it contains. Note it and remove its value from the integer. Proceed recursively to find the unique representation. E.g., for $n = 4$, $F_4 = 3$ is the largest Fibonacci number that is contained in 4 and $F_2 = 1$ is the largest Fibonanni number that is contained in $n - F_4 = 1$. This gives us the representation 1011.

Recall that besides a recursive description of the Fibonacci numbers there exists an explicit formula $F_i = \lfloor \frac{\phi^i}{\sqrt{5}} + \frac{1}{2} \rfloor$, where $\phi = \frac{1+\sqrt{5}}{2} \sim 1.618$ is the golden ratio.

(i) [5pts] What is the length of $\mathcal{C}_F(n)$?

(ii) [2pts] Show that the code is prefix-free. *Hint*: Use the property of Fibonacci numbers.

(iii) [3pts] Show that $\log_\phi(\sqrt{5}i) \leq 3 + 2\log_2 i$.

(iv) [5pts] Consider a random variable U that takes values on the positive integers s.t. $P(U = i)$ is decreasing. Show that $\mathbb{E}[\text{length}(\mathcal{C}_F(U))] \leq 3 + 2H(U)$. *Hint*: First show that $iP(U = i) \leq 1$.

(space for problem 3)

(space for problem 3)

Problem 4. *(Exponential families - 10pts)*

For $t > 0$, consider a family of distributions supported on $[t, +\infty]$ such that $\mathbb{E}[\ln X] = \frac{1}{\alpha} + \ln t$, $\alpha > 0$.

(i) [5pts] What is the parametric form of a maximum entropy distribution satisfying the constraint on the support and the mean?

(ii) [5pts] Find the exact form of the distribution.

(space for problem 4)

(space for problem 4)

Problem 5. *(KL Divergence and L1 – 10pts)*

(i) [5pts] Show that $-D(p||q) \leq \log \sum_i \min(p_i, q_i) + \log \sum_i \max(p_i, q_i)$ *Hint:* The key is to write $\log(q/p)$ in some clever form involving min and max. And if you do not know what to do with sums involving logs, ask Jensen.

(ii) [5pts] Use the fact that $\min(a, b) = \frac{a+b}{2} - \frac{|a-b|}{2}$ and $\max(a, b) = \frac{a+b}{2} + \frac{|a-b|}{2}$ and (a) to show that $|p - q|_1 \leq 2\sqrt{1 - \exp(-D(p||q))}$.

(space for problem 5)

(space for problem 5)

Problem 6. *(Signal Representations – 10 pts)* Assume that we get m samples in $\mathbb{R}^d$, call them $u_1, \ldots, u_m$. The dimension d is very large. Therefore we would like to compress the data. We fix $n < d$ and we would like to produce n-dimensional representations $\hat{u}_1, \ldots, \hat{u}_m$ that are close to the original ones. Assume that we collect our data samples into a $d \times m$ matrix U and the desired representations into a $n \times m$ matrix $\hat{U}$.

In the course we learned that two possible compression techniques for this scenario are the PCA and random projections.

Recall that random projections are linear maps $f(u) : \mathbb{R}^d \to \mathbb{R}^n$, defined as $f(u) = \frac{1}{\sqrt{n}} X u$, where X is a real-valued matrix with iid zero-mean unit-variance entries.

(i) [5pts] Assume that your "goodness" criterion is the spectral norm $||U^T U - \hat{U}^T \hat{U}||_2$. What guarantees to you get for both methods? You can assume that the smallest eigenvalue of $X^T X$ is 0.

(ii) [5pts] Assume your "goodness" criterion is $\max_{i,j} |||u_i - u_j||^2 - ||\hat{u}_i - \hat{u}_j||^2|$. What guarantees do you get for both methods? No need for complicated computations.

(space for problem 6)

(space for problem 6)

(space for whatever you want to write here)