Advanced Topics in Machine Learning 2024

Amartya Sanyal

Home Assignment 4

Deadline: 23:59, Thursday, 10 October 2024

*The assignments must be submitted individually – each student must write and submit a personal solution but we do not prevent you from discussing **high-level** ideas in small groups. If you use any LLM tool such as ChatGPT, please specify the **purpose** and **manner** in which you have utilized it.*

We are interested in how *you solved the problems, and not just in the final answers. Please explain your solutions and ideas as clearly as you can.*

Late Penalty and multiple Submissions *Late submissions will incur a penalty of 10% of the total marks for every hour of delay (rounded up) with a maximum allowed delay of 5 hours after which the submission server will close. If you submit multiple submissions, only the last submission will be considered relevant both for grading answers as well as late penalty.*

Submission format: *Please upload your answers in a single* `.pdf` *file. If you have created code please include at least the key parts in the PDF as well as a link to the full code (on Github, Colab, or similar).*

1. We have studied the *privacy loss random variable* in class. For a randomized algorithm $\mathcal{A}$ and neighboring datasets D and D', the privacy loss random variable is defined as:

$$\text{PrivLoss}\left(\mathcal{A}\left(D\right)||\mathcal{A}\left(D'\right)\right) = \ln\left(\frac{\Pr[\mathcal{A}(D) = o]}{\Pr[\mathcal{A}(D') = o]}\right),$$

where o is distributed as $\mathcal{A}(D$.

Below (in Q2), we will study the probability density functions (pdfs) of different privacy loss random variables. Each pdf corresponds to one or more of the following variants of differential privacy for certain ranges of the relevant parameters $(\epsilon, \delta, \rho, \alpha, \epsilon'(\alpha))$:

(1) Pure Differential Privacy (ϵ-DP)

(2) Approximate Differential Privacy (ϵ, δ-DP)

(3) Zero-Concentrated Differential Privacy (ρ-zCDP)

(4) Rényi Differential Privacy ($\alpha, \epsilon'(\alpha)$-RDP)

Question 1 (25 points)**.** *Clearly state and rigorously justify the relationships between the above notions of differential privacy.*

Your answer should indicate for each pair of the above notions whether one implies the other or not, and if so, which one implies the other. Include a clear argument (no more than five sentences per claim) using the privacy loss random variable to support your statements.
Hint: *For most pairs, use the argument that 'a' implies 'b' and 'b' implies 'c'; so 'a' implies 'c'. You do not need to state the exact privacy parameters with which one implies the other—for example, e.g. if pure DP with parameter ϵ implies approximate DP with parameters (ϵ', δ'), you do not need to state the relationship between ϵ, ϵ', and δ'.*

Following was released mistakenly in a previous update of this assignment on absalon, so consider them as hints in solving the question: "Approx DP is the least restrictive, followed by Renyi DP (sub-exponential privLoss), followed by z-CDP (sub-gaussian privLoss), followed by pure (bounded privLoss)"

Answer:

Next, we will study the following five probability density functions (pdfs) of five different privacy loss random variables corresponding to the above notions of differential privacy. Assume that each pdf holds for all pairs of neighboring datasets D, D' for the respective mechanism.

(a) A delta function at $L = \epsilon_0$, i.e., $f(L) = \delta(L - \epsilon_0)$.

(b) A uniform distribution over $L \in [0, \epsilon_1]$, i.e., $f(L) = \frac{1}{\epsilon_1}$ for $0 \leq L \leq \epsilon_1$.

(c) A normal (Gaussian) distribution centered at $L = 0$ with variance σ^2, i.e.,

$$f(L) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-L^2/(2\sigma^2)}.$$

(d) A Laplace distribution centered at $L = 0$ with scale parameter b, i.e.,

$$f(L) = \frac{1}{2b} e^{-|L|/b}.$$

(e) A mixture of two delta functions at $L = \epsilon_2$ and $L = -\epsilon_2$, each with probability 0.5, i.e.,

$$f(L) = \frac{1}{2}\delta(L - \epsilon_2) + \frac{1}{2}\delta(L + \epsilon_2).$$

Question 2 (25 points)**.** *For each of the Privacy Loss Random Variables given above, state **all** differential privacy mechanism(s) from the list in Question 1 that it represents. Provide mathematically rigorous justifications for each, keeping your explanation to a maximum of three sentences per Privacy Loss Random Variable.*
Hint: *Only prove for the strictest definition and use the implications from Question 1.*

Following was released mistakenly in a previous update of this assignment on absalon, so consider them as hints in solving the question: "a,b,e are pure DP. c is z-CDP and d is RenyiDP"

2. Ignore this question as the answers have been mistakenly released. Go to question 5 and 6 next. But feel free to read the answers of Question 3 and 4. Next, we will prove that the VC dimension of linear halfspaces in $\mathbb{R}^n$ is $n + 1$. A linear halfspace $h_{\mathbf{w}}$ in $\mathbb{R}^n$ is defined by the parameter vector $\mathbf{w} = (w_0, w_1, \ldots, w_n)$ and classifies any $x \in \mathbb{R}^n$ as follows:

$$h_{\mathbf{w}}(x) = \mathbb{I}\left\{\sum_{i=1}^{n} w_i x_i + w_0 \geq 0\right\}.$$

Question 3 (0 points)**.** *Construct a set of $n+1$ points in $\mathbb{R}^n$ that is shattered by the class of linear halfspaces in $\mathbb{R}^n$.*

Answer: Cosnider the zero vector, and the n canonical vectors. Then to shatter any labelling, set $w_0 = 1/2$ and w_i to be +1 if the label of the i^{th} canonical vector is +1 and -1 otherwise.

For the next question, we will need an additional theorem. Given a set $P = \{x_1, \ldots, x_k\} \subset \mathbb{R}^n$, define its convex hull as

$$\left\{z \in \mathbb{R}^n \mid \exists \lambda_1, \ldots, \lambda_k, \sum_{i=1}^{k} \lambda_i = 1 \wedge z = \sum_{i=1}^{k} \lambda_i x_i\right\}.$$

Radon's theorem states that if $k \geq n + 2$, then P must have two disjoint subsets P_1 and P_2 whose convex hulls intersect.

Question 4 (0 points)**.** *Using the theorem above, prove that no set of size $n + 2$ can be shattered by the hypothesis class of linear halfspaces in $\mathbb{R}^n$.*
Hint: *Assume such a set exists. Apply Radon's theorem to that set and use the definition of the convex hull to arrive at a contradiction.*

Answer: Consider any set S of n+ 2 points. Let S_1 and S_2 be the disjoint subsets of S obtained by an application of Radon's theorem. Let all points in S1 be labelled 1 and all points in S2 be labelled 0, which is a valid labelling since S1 and S2 are disjoint. Suppose there exists a linear halfspace, defined by $\mathbf{w} \in \mathbb{R}^n, w_0 \in \mathbb{R}$, such that $\langle w, x \rangle + w_0 \geq 0$ for all $x \in S_1$ and $\langle w, x \rangle + w_0 \leq 0$ for all $x \in S_2$. Then, it also must be the case that any z in the convex hull of S_1 satisfies $\langle w, x \rangle + w_0 \geq 0$ and any z in the convex hull of S2 satisfies $\langle w, x \rangle + w_0 \leq 0$. But according to Radon's theorem these two convex hulls intersect, a ontradiction. Therefore, there can exist no linear halfspace consistent with this particular labelling.

3. Next, we will prove that the VC dimension of convex sets in the unit square $[0,1]^2$ is ∞. Let $S \subset [0,1]^2$ be any convex set in the unit square $[0,1]^2$ and let $c_S : [0,1] \rightarrow \{0,1\}$ be a classifier where $c_S(x) = 1$ if $x \in S$ and 0 otherwise. Let $\mathcal{C} = \left\{ c_S \middle| \; S \text{ is a convex set in } [0,1]^2 \right\}$.

   Question 5 (35 points)**.** *Prove that the VC dimension of $\mathcal{C}$ is ∞.*

   Question 6 (15 points)**.** *What does this imply about PAC learnability of $\mathcal{C}$. Write no more than two sentences.*