۱	۴	۷	
۲	۵	جمع نمرات	
۳	۶		

دانشگاه صنعتی شریف

شماره دانشجویی:

نام و نام خانوادگی:

شماره صندلی: نام استاد:

شماره سوال: سوال ۱ : اضافه کردن نویز لاپلاس با مقیاس $\frac{1}{\varepsilon}$ ویژگی ε-DP یا همان $O(\varepsilon)$ محرمانه تفاضلی را دارد. ولی اضافه کردن نویز یکنواخت از بازه $[-\frac{3}{\varepsilon}, \frac{3}{\varepsilon}]$ ویژگی ε-DP را ندارد و $O(\varepsilon)$-محرمانه تفاضلی نیست. **جواب: خیر** در ادامه علت را توضیح می دهیم.

چرا نویز لاپلاس ε-DP است؟ در اسلاید ۱۱ از Lec19 این اثبات آمده است در اینجا به اختصار آن را بیان می کنیم.

دو دیتاست همسایه که در یک record متفاوت هستند در count شان حداکثر ۱ واحد متفاوت هستند.

فرض کنید s و s' در مجموع ممکن به جواب پرسش باشند که از دیتاست های همسایه هستند پس $|s-s'|\leq 1$

چگالی توزیع لاپلاس در نقطه y برابر است با:

$$Lap_b(y) = \frac{1}{2b}\exp\left(\frac{-|y|}{b}\right)$$

فرض کنید $b=\frac{1}{\varepsilon}$ پس چگالی خروجی وقتی مجموع واقعی s هست برابر است با

$$P_s(y) = \frac{\varepsilon}{2}e^{-\varepsilon|y-s|}$$

پس برای همسایه های s و s' با فرض $|s-s'|\leq 1$ داریم:

$$\frac{P_{s'}(y)}{P_s(y)} = \exp(\varepsilon(|y-s|-|y-s'|))$$

برای $y\in\mathbb{R}$ و اعداد a و b با شرط $|a-b|\leq 1$ (در اینجا $a=s$ و $b=s'$) نامساوی مثلث بیان می کند که

$$|y-a|-|y-b|\leq|a-b|\leq 1 \Longrightarrow |y-s|-|y-s'|\leq|s-s'|\leq 1$$

پس

$$\frac{P_{s'}(y)}{P_s(y)} \leq \exp(\varepsilon(|y-s|-|y-s'|)) \leq \exp(\varepsilon\times 1) = \exp(\varepsilon) \longrightarrow \boxed{\frac{P_{s'}(y)}{P_s(y)}\leq e^{\varepsilon}}$$

به طور مشابه نسبت معکوس هم این ویژگی را دارد یعنی $\frac{P_s(y)}{P_{s'}(y)}\leq e^{\varepsilon}$ بنابراین برای هر مجموعه قابل اندازه گیری S (measurable) داریم:

$$Pr[\tilde{f}(x)\in S|s'] \leq e^{\varepsilon} Pr[\tilde{f}(x)\in S|s]$$

بنابراین مکانیزم لاپلاس با پارامتر $\frac{1}{\varepsilon}$ دارای ε-DP است.

CamScanner

★ غیرممکن است چراکه support های آن ها shift می خورد. بنابراین برای این خروجی ها نسبت احتمال نامحدود و ∞ می شود بنابراین نمی توان ε-DP خالص داشت و توزیع یکنواخت $O(\varepsilon)$ محرمانه تفاضلی نیست.

چرا نویز یکنواخت در بازه $\left[\frac{-3}{\varepsilon}, \frac{3}{\varepsilon}\right]$ ε-DP نیست و $O(\varepsilon)$ محرمانه تفاضلی نیست؟ با مثال نقض اثبات می کنیم.

فرض کنید $Z \sim Uniform\left[\frac{-3}{\varepsilon}, \frac{3}{\varepsilon}\right]$ برای یک مجموع واقعی s چگالی احتمال خروجی می شود:

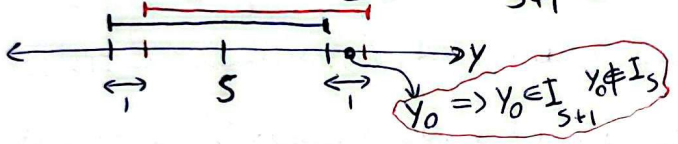
$$u_s(y) = \begin{cases} \frac{\varepsilon}{6}, & y \in \left[s - \frac{3}{\varepsilon}, s + \frac{3}{\varepsilon}\right] \\ 0, & \text{otherwise} \end{cases}$$

در مجموع همسایه s و $s+1$ را در نظر بگیرید. (مثلاً در یک dataset داریم $X_j = 0$ و در دیگری $X_j = 1$)

support های آنها بازه های زیر هستند:

$$I_s = \left[s - \frac{3}{\varepsilon}, s + \frac{3}{\varepsilon}\right], \quad I_{s+1} = \left[s+1 - \frac{3}{\varepsilon}, s+1 + \frac{3}{\varepsilon}\right]$$

توجه کنید این دو بازه نسبت به هم یک واحد شفیت خورده اند.



اجتماع این دو support دارای طول $\frac{6}{\varepsilon} + 1$ است.

$$|I_s \cup I_{s+1}| = |I_s| + |I_{s+1}| - |I_s \cap I_{s+1}| = \frac{6}{\varepsilon} + \frac{6}{\varepsilon} - \left(\frac{6}{\varepsilon} - 1\right) = \frac{6}{\varepsilon} + 1$$

بنابراین همیشه یک ناحیه وجود دارد که نقاط در آن ناحیه در یک support وجود دارند و در دیگری وجود ندارند.

به طور مشخص برای مثال ما مقدار زیر را انتخاب می کنیم:

$$y_0 = s + 1 + \frac{3}{\varepsilon} - \delta, \quad \delta \in (0,1)$$

برای یک $0 < \delta < 1$ داریم: $y_0 \in I_{s+1}$ ولی $y_0 \notin I_s$ بنابراین:

$$u_{s+1}(y_0) = \frac{\varepsilon}{6} > 0, \quad u_s(y_0) = 0$$

پس نسبت likelihood برابر است با

$$\frac{u_{s+1}(y_0)}{u_s(y_0)} = \infty$$

که هر bound حاصل از ضرب عددی در $\exp(\varepsilon)$ را نقض می کند. بنابراین یک measurable set (هر همسایگی کوچکی از y_0) وجود دارد که $\Pr[\tilde{f} \in S \mid s+1] > 0$ ولی $\Pr[\tilde{f} \in S \mid s] = 0$ یک با ε-DP تناقض دارد پس توزیع یکنواخت $O(\varepsilon)$ محرمانه تفاضلی نیست.

شهود: ε-DP خالص نیاز دارد که برای هر خروجی ممکن y چگالی های شرطی احتمال مثبت باشند و نسبت آنها با $\exp(\varepsilon)$ محدود باشد. توزیع لاپلاس دارای full support در $\mathbb{R}$ است (همه جا چگالی مثبت دارد). بنابراین نسبت احتمال محدود است و دارای bound با ضریبی از $\exp(\varepsilon)$ است. یک توزیع bounded-support مانند توزیع یکنواخت مقادیر خروجی ای دارد که برای یک دیتاست همسایه ممکن است ولی برای دیتاست مجاور آن

★ ادامه بالای صفحه

CamScanner

سوال ۲ ۱. رویکرد برش دهی ساده

اجازه دهید $T = R + C\sqrt{\log(n)}$ با $C > 0$ یک عدد ثابت به اندازه کافی بزرگ. بنابراین تخمین گر برش خورده می شود:

$$\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \cdot 1_{[|X_i| \le T]}$$

پس ℓ_1-sensitivity تخمین گر $\tilde{\mu}$ برابر است با $\Delta = \frac{2T}{n}$ چرا که تغییر دادن یک X_i خاص جمع را حداکثر می تواند $2T$ تغییر دهد. (دلیل: هر عبارت دارای کران T برای قدر مطلق خود است. پس max تفاوت برابر $T-(-T)=2T$ است و در حالت extreme داریم $X_i = T$ و $X_i' = -T$ چون $-T \le X_i \cdot 1_{[|X_i| \le T]} \le T$)

بنابراین خروجی الگوریتم A برابر است با:

$$A(X) = \tilde{\mu} + Lap\left(\frac{\Delta}{\varepsilon}\right) \Rightarrow \boxed{A(X) = \tilde{\mu} + Lap\left(\frac{2T}{n\varepsilon}\right)}$$

که این الگوریتم $(\varepsilon, 0)$-DP است بوسیله مکانیزم لاپلاس.

حال به سراغ تحلیل خطای الگوریتم می رویم: توجه کنید خطای الگوریتم می شود: $|A(X) - \mu|$

می نویسیم $A(X) - \mu = (\tilde{\mu} - \mu) + Lap\left(\frac{2T}{n\varepsilon}\right)$

حال $|\tilde{\mu} - \mu|$ را با احتمال بالا bound می کنیم. فرض کنید متغیر تصادفی Y_i می شود: $Y_i = X_i \cdot 1_{[|X_i| \le T]}$

پس:

$$\tilde{\mu} - \mu = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)1_{[|X_i| \le T]}}_{\text{عبارت اول}} - \underbrace{\frac{\mu}{n}\sum_{i=1}^{n} 1_{[|X_i| > T]}}_{\text{عبارت دوم}}$$

حال چون $X_i \sim N(\mu, 1)$ و $\mu \in [-R, R]$ داریم:

$$\Pr[|X_i| > T] \le \Pr[|X_i - \mu| > C\sqrt{\log(n)}] \le \underbrace{2\exp\left(-\frac{C^2\log(n)}{2}\right)}_{2n^{-\frac{C^2\log(e)}{2}} = 2n^{-\frac{C^2}{2}}} = 2n^{-\frac{C^2}{2}}$$

حال با استفاده از یک Chernoff bound با احتمال $1 - e^{-\Omega(n)}$ تعداد نقاط برش داده شده حداکثر $O(n^{1-\frac{C^2}{2}})$ است.

بنابراین عبارت دوم برای یک C به اندازه کافی بزرگ $O(n^{-\frac{C^2}{2}}) = O(\sqrt{\frac{1}{n}})$. همچنین برای عبارت اول توجه کنید که $(X_i - \mu)1_{[|X_i| \le T]}$ با $O(T)$ کران دار است و همچنین variance ≤ 1 چون $X_i \sim N(\mu, 1)$

پس بوسیله Bernstein's inequality با احتمال 0.995 داریم: $\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)1_{[|X_i| \le T]}\right| \lesssim \sqrt{\frac{1}{n}}$

حال چون برای هر دو عبارت اول و عبارت دوم $O(\sqrt{\frac{1}{n}})$ را نشان دادیم پس:

$$|\tilde{\mu} - \mu| \lesssim \sqrt{\frac{1}{n}} \quad \text{with probability} \ge 0.99$$

CamScanner

* مثلاً لاپلاس با مقیاس $\frac{1}{\varepsilon}$ تا زمانی که n در مقایسه با نویز بزرگ باشد argmax را عوض نخواهد کرد. که این همان ایده پشت پیدا کردن high-count bin به شکل private است.

عبارت نویز لاپلاس $\left(Lap\left(\frac{2T}{n\varepsilon}\right)\right)$ با احتمال $1-\delta$ برابر $O\left(\frac{T\log(\frac{1}{\delta})}{n\varepsilon}\right)$ است.

مثلاً $\underbrace{\delta=0.01}_{probability=0.99}$ می‌گیریم. بنابراین با احتمال حداقل 0.97 (probability ≥ 0.97) خواهیم داشت

$$|A(x_1,\dots,x_n)-\mu| \lesssim \sqrt{\frac{1}{n}}+\frac{T}{n\varepsilon} \xrightarrow{T=R+C\sqrt{\log(n)}} |A(x)-\mu| \lesssim \sqrt{\frac{1}{n}}+\frac{R+C\sqrt{\log(n)}}{n\varepsilon}$$

حکم ثابت شد.

۲. روش هیستوگرام: بازه $[-R,R]$ را به بازه‌های به طول واحد I_j افرازی کنیم که j عدد صحیح است. فرض کنید f_i تعداد نمونه‌های موجود در سطل I_j باشد. همچنین فرض کنید j^* اندیسی است که f_j را بیشینه می‌کند.

ابتدا سطل (بازه) ای که شامل مکان μ است را ثابت می‌کنیم. فرض کنید j_0 عدد صحیحی است که مربوط به بازه‌ای است که شامل μ است. احتمال اینکه یک نمونه از توزیع گاوسی $(x_i \sim N(\mu,1))$ در یک بازه با طول واحد که μ در مرکز آن باشد برابر است با:

$$p_0=\Pr\left[|x_i-\mu|\leq\frac{1}{2}\right]=\Phi\left(\frac{1}{2}\right)-\Phi\left(\frac{-1}{2}\right)\approx 0.3829$$

فرض $\boxed{\mu\in I_{j_0}}$

بنابراین تعداد مورد انتظار (expected) در این سطل (بازه به طول واحد) برابر است با $p_0 n=0.3829n$

حال بوسیله یک Chernoff bound با احتمال $1-\exp(-\Omega(n))$ داریم:

$$f_{j_0}\geq p_0 n-O(n)\geq 0.3\,n\geq 0.1n \Rightarrow \boxed{f_j(x_1,\dots,x_n)\geq 0.1n}$$

حال هر سطل I_j را که فاصله مرکز آن از μ بیشتر از 5 است را در نظر بگیرید. $(|\mu-j|>5)$ بنابراین برای هر نقطه در این بازه (سطل) داریم $|x-\mu|\geq 4.5$ (چرا که بازه‌ها به طول واحد اند و فرض کردیم μ در وسط بازه است.) پس احتمال اینکه یک نمونه در این بازه بیفتد حداکثر برابر است با:

$$\underbrace{p_{far}}\leq \Pr[|X-\mu|\geq 4.5]\leq \exp\left(\frac{-4.5^2}{2}\right)=\underbrace{\exp(-10.125)}_{\simeq 4\times 10^{-5}}\ll 0.02 \rightarrow \boxed{p_{far}<0.02}$$

بنابراین تعداد مورد انتظار (expected count) در یک سطل (بازه) دور ناچیز است و برابر $p_{far}\,n<0.02n$ است و با استفاده از Chernoff bound با احتمال $1-\exp(-\Omega(n))$ تعداد واقعی موجود در سطل‌های دورتر ناچیز و $<0.02n$ است. بنابراین هیچ یک از سطل‌های I_j دورتر نمی‌توانند سطل مرکزی I_0 که شامل μ است را شکست دهند و بنابراین سطل با دارای بیشترین تعداد j^* (max count bin) باید در فاصله حداکثر 5 از میانگین قرار بگیرد یعنی $|\mu-j^*|\leq 5$ با احتمال زیاد $\geq 1-\exp(-\Omega(n))$ و بنابراین bound های گفته شده در سوال ثابت می‌شوند.

اگر از یک الگوریتم هیستوگرام محرمانه بجای مقدارها واقعی (true counts) استفاده کنیم فاصله زیاد (large margin) موجود یعنی $f_{j_0}-\max_{|j-\mu|>5} f_j\geq 0.3n$ اطمینان حاصل می‌کند که privacy noise اضافه شده که کوچک است

* ادامه بالای صفحه

$$|A(x)-\mu| \lesssim \sqrt{\frac{1}{n}} + \frac{R+\sqrt{\log(n)}}{n\varepsilon} \xrightarrow[n \gtrsim \frac{\log(R)}{\varepsilon}]{\text{روش هیستوگرام}} |A(X)-\mu| \lesssim \sqrt{\frac{1}{n}} + \frac{O(1)+\sqrt{\log(n)}}{n\varepsilon} \quad *$$

۱	۴	۷
۲	۵	جمع نمرات
۳	۶	

دانشگاه صنعتی شریف

شماره دانشجویی:

نام و نام خانوادگی:

شماره صندلی: نام استاد:

شماره سوال: ۳. ترکیب تطبیقی و کران های بهبود یافته

وجود الگوریتم را با معرفی آن نشان می دهیم. الگوریتم کارآمد DP-$(\varepsilon, 0)$ به شکل زیر است:

الگوریتم A شامل ۲ مرحله است:

۱. از یک الگوریتم هیستوگرام (محرمانه) روی بازه (سطل) های I_j به طول واحد استفاده می کنیم تا یک j^* که بازه با بیشترین تعداد (محرمانه) را پیدا کنیم. طبق قسمت (۲) همین سوال با احتمال زیاد j^* شرط $|\mu - j^*| \leq 5$ را برآورده می کند.

چون می خواهیم 5 تا بازه قبل و بعد را شامل شود 6 در نظر می گیریم.

۲. حال توجه را به بازهٔ $I=[j^*-6, j^*+6]$ دارای شعاع ثابت (مثلاً در اینجا شعاع = 6) محدود می کنیم.

هر X_i را برش می دهیم تا در این بازهٔ کوچک به شعاع ثابت قرار گیرد. (به علاوه عبارت $C\sqrt{\log(n)}$)

سپس الگوریتم برش دهی ساده برای تخمین میانگین توزیع گاوسی + مکانیزم لاپلاس در قسمت (۱) همین سوال ولی با کران $\mu \in [j^*-6, j^*+6]$ اجرا می کنیم در واقع با $T' = \underbrace{O(1)}_{\text{عدد ثابت}} + C\sqrt{\log(n)}$ به جای استفاده از $T = R + C\sqrt{\log(n)}$ الگوریتم را اجرا می کنیم.

چون شعاع ثابت است مثلاً در اینجا

تحلیل خطای الگوریتم: با احتمال حداقل $1-\exp(-\Omega(n))$ مرحله ۱ (هیستوگرام) یک سطل (بازه) با حداکثر فاصله 5 از μ را مشخص می کند. وقتی که شعاع ثابت این بازه کوچک دانسته شود مثلاً 5 در اینجا آنگاه bound برش T' برابر است با $O(1)+C\sqrt{\log(n)}$ که مستقل از R است. سپس با استفاده از پارامتر sensitivity/noise دقیقاً مشابه قسمت (۱) همین سوال پارامتر مقیاس مکانیزم لاپلاس می شود $\frac{2T'}{n\varepsilon} = O\left(\frac{1+\sqrt{\log(n)}}{n\varepsilon}\right)$ و خطای نمونه برداری همان $O\left(\sqrt{\frac{1}{n}}\right)$ باقی می ماند. اگر n به اندازه کافی بزرگ باشد بنابراین هر دوی $\sqrt{\frac{1}{n}}$ و $\frac{1+\sqrt{\log(n)}}{n\varepsilon}$ خیلی کوچک تر از 5 $(\ll 5)$ می شوند و خروجی حداکثر در فاصلهٔ 5 از μ با احتمال بالای ادعا شده خواهد بود. به طور مشخص، اگر $n \gtrsim \frac{\log(R)}{\varepsilon}$ (پس مرحله ۱ یا همان مرحله هیستوگرام به طور محرمانه موفق خواهد بود و count gap بر نویز محرمانه غلبه می کند.) می توانیم ضمانت کنیم با احتمال حداقل $0.99-\exp(-\Omega(n))$ داریم $|A(x_1,\dots,x_n)-\mu| \leq 5$

چرا این نتیجه نشان دهنده یک بهبود نمایی در وابستگی به R است؟ در قسمت (۱) این سوال عبارت نویز جمع شونده یک فاکتور وابسته R داشت $(T \approx R)$ که منجر به عبارت خطای $\propto \frac{R}{n\varepsilon}$ می شد. این عبارت رابطه خطی با R دارد. در روش تطبیقی ما ابتدا μ را localize می کنیم در یک شعاع ثابت با روش هیستوگرام

بالای صفحه را هم ببینید. *

CamScanner

بنابراین وابستگی به R از عبارت sensitivity/noise حذف می‌شود و الگوریتم فقط وابسته به constant ها و n که $O(1)$ می‌شود. همچنین مرحله (۱) یا همان هیستوگرام فقط به اندازه $\frac{\log R}{\varepsilon}$ نمونه نیاز دارد تا بازه (سطل) نزدیک M و محرمانه را پیدا کند. بنابراین وابستگی به R تبدیل به فقط وابستگی به R $O(\log(R))$ می‌شود.

که به معنای بهبود نمایی در مقایسه با وابستگی خطی R است.

سوال ۳ الف) چون قضیه اگر و تنها اگر است باید دو جهت آن را ثابت کنیم.

جهت ۱: $(\varepsilon,\delta)\text{-DP} \Rightarrow$ Hockey-stick bound

فرض کنید M یک مکانیزم (ε,δ)-DP است. پس برای تمامی measurable set هایی نظیر S داریم:

$$\Pr[M(x')\in S]\leq e^{\varepsilon}.\Pr[M(x)\in S]+\delta$$

حال مجموعه S^* را به این شکل تعریف می‌کنیم: $S^*=\{y: f_{x'}(y)>e^{\varepsilon}f_x(y)\}$

این مجموعه‌ای است که x' احتمال خیلی زیادی نسبت به x دارد. برای این مجموعه همانند تمامی مجموعه‌ها داریم:

$$\Pr[M(x')\in S^*]\leq e^{\varepsilon}\Pr[M(x)\in S^*]+\delta$$

$$\int_{S^*} f_{x'}(y)\,dy \leq e^{\varepsilon}.\int_{S^*} f_x(y)\,dy+\delta \Rightarrow \int_{S^*}[f_{x'}(y)-e^{\varepsilon}f_x(y)]\,dy\leq\delta$$

همچنین توجه کنید که

$$H_{e^{\varepsilon}}(M(x')\|M(x))=\int[f_{x'}(y)-e^{\varepsilon}f_x(y)]_+\,dy$$

بنابراین فقط تفاوت در $[\cdot]_+$ برای اثبات است. توجه کنید که بخش مثبت $[\cdot]_+$ یعنی ما فقط باید انتگرال را روی y هایی که $f_{x'}(y)>e^{\varepsilon}f_x(y)$ هست بگیریم که دقیقاً همان تعریف S^* است پس داریم:

$$\int_{S^*}[f_{x'}(y)-e^{\varepsilon}f_x(y)]\,dy=\int_{S^*}[f_{x'}(y)-e^{\varepsilon}f_x(y)]_+\,dy$$

بنابراین حکم ثابت می‌شود و داریم:

$$H_{e^{\varepsilon}}(M(x')\|M(x))=\int_{S^*}[f_{x'}(y)-e^{\varepsilon}f_x(y)]_+\,dy\leq\delta$$

جهت اول ثابت شد.

توجه کنید که S^* بدترین حالت بود و برای measurable set های دیگر که شامل اعضایی با شرط $y: f_{x'}(y)\leq e^{\varepsilon}f_x(y)$ روابط بالا برقرار است چرا که وجود چنین y هایی باعث کاهش یافتن $H_{e^{\varepsilon}}(M(x')\|M(x))$ می‌شود چرا که در عملگر $[\cdot]_+$ صفر می‌شود و margin میان $H_{e^{\varepsilon}}(M(x')\|M(x))$ با δ بیشتر خواهد شد.

CamScanner

حال جهت دوم را اثبات می‌کنیم: Hockey-stick bound $\Longrightarrow (\varepsilon,\delta)$-DP

فرض کنید برای هر دو مجموعه داده مجاور X و X' داریم: $H_{e^\varepsilon}(M(X')\|M(X)) \le \delta$

برای هر measurable set نظیر S داریم: $\Pr[M(X')\in S]=\int_S f_{X'}(y)\,dy$

حال S را به دو بخش افرازی کنیم:

$$S_1 = S\cap\{y: f_{X'}(y)\le e^{\varepsilon}\cdot f_X(y)\}$$
$$S_2 = S\cap\{y: f_{X'}(y)> e^{\varepsilon}\cdot f_X(y)\}$$

برای S_1 داریم: $\boxed{\int_{S_1} f_{X'}(y)\,dy \le e^{\varepsilon}\cdot\int_{S_1} f_X(y)\,dy}$ (*)

برای S_2 داریم:

$$\int_{S_2} f_{X'}(y)\,dy = \int_{S_2}[f_{X'}(y)-e^{\varepsilon}f_X(y)]\,dy + e^{\varepsilon}\int_{S_2} f_X(y)\,dy$$

$$\rightarrow \boxed{\int_{S_2} f_{X'}(y)\,dy \le H_{e^\varepsilon}(M(X')\|M(X)) + e^{\varepsilon}\int_{S_2} f_X(y)\,dy \le \delta + e^{\varepsilon}\int_{S_2} f_X(y)\,dy} \quad (**)$$

از فرض سوال استفاده کردیم

با ترکیب کردن عبارات S_1 و S_2 داریم:

$$\Pr[M(X')\in S] = \int_{S_1} f_{X'}(y)\,dy + \int_{S_2} f_{X'}(y)\,dy \overset{(*)(**)}{\le} e^{\varepsilon}\cdot\int_{S_1} f_X(y)\,dy + \delta + e^{\varepsilon}\int_{S_2} f_X(y)\,dy$$

$$\Pr[M(X')\in S] \le e^{\varepsilon}\cdot\underbrace{\left[\int_{S_1} f_X(y)\,dy + \int_{S_2} f_X(y)\,dy\right]}_{\int_S f_X(y)\,dy} + \delta = e^{\varepsilon}\cdot\Pr[M(X)\in S] + \delta$$

$$\Rightarrow \boxed{\Pr[M(X')\in S] \le e^{\varepsilon}\cdot\Pr[M(X)\in S] + \delta}$$

بوسیله تقارن یا همان symmetry (با جابه‌جایی X و X') ما همچنین نامساوی برعکس را نتیجه می‌گیریم که اثبات را کامل می‌کند.

ب) فرض کنید X و X' دو مجموعه داده مجاورانند. برای هر measurable set همچون S داریم:

$$\Pr[M(X')\in S] = \int_S f_{X'}(y)\,dy$$

حال S را با توجه به $\overbrace{\text{privacy loss}}^{\text{اتلاف محرمانه}}$ به دو مجموعه افرازی می‌کنیم:

$$S_{low}=\{y\in S: L_{X'\|X}(y)<\varepsilon\}=\{y\in S: f_{X'}(y)<e^{\varepsilon}.f_X(y)\}$$

$$S_{high}=\{y\in S: L_{X'\|X}(y)\geq\varepsilon\}=\{y\in S: f_{X'}(y)\geq e^{\varepsilon}.f_X(y)\}$$

$$S=S_{high}\cup S_{low}$$

برای S_{low} داریم:

$$\boxed{\int_{S_{low}} f_{X'}(y)dy<\int_{S_{low}} e^{\varepsilon}.f_X(y)\,dy=e^{\varepsilon}\int_{S_{low}} f_X(y)dy} \quad (*)$$

برای S_{high} داریم:

$$\boxed{\int_{S_{high}} f_{X'}(y)dy\leq\int_{\text{all } y:L_{X'\|X}(y)\geq\varepsilon} f_{X'}(y)dy=\Pr[L_{X'\|X}(Y')\geq\varepsilon]\leq\delta} \quad (**)$$

چرا که $S_{high}\subseteq\{y:L_{X'\|X}(y)\geq\varepsilon\}$

که در آن $Y'\sim M(X')$ است. توجه کنید در نامساوی آخر از فرض سوال استفاده شد.

حال رابطهٔ $(*)$ و $(**)$ را ترکیب می‌کنیم:

$$\Pr[M(X')\in S]=\int_S f_{X'}(y)dy=\int_{S_{low}} f_{X'}(y)dy+\int_{S_{high}} f_{X'}(y)dy\underset{(*)(**)}{\leq}e^{\varepsilon}.\int_{S_{low}} f_X(y)dy+\delta$$

$$\xrightarrow{S_{low}\subseteq S}\boxed{\Pr[M(X')\in S]\leq e^{\varepsilon}.\int_S f_X(y)dy+\delta=e^{\varepsilon}.\Pr[M(X)\in S]+\delta}$$

$$\boxed{\Pr[M(X')\in S]\leq e^{\varepsilon}.\Pr[M(X)\in S]+\delta}$$

توجه کنید چون ترتیبی برای اثبات نداشتیم بوسیله تقارن یا symmetry با جابه‌جایی X و X' نامساوی برعکس نیست ثابت می‌شود. پس مکانیزم M در تعریف (ε,δ)-DP صدق می‌کند.

ج) خیر. شرط tail bound موجود در قسمت (ب) دقیق (tight) نیست. در واقع شرط کافی است ولی شرط لازم نیست. یعنی (ε,δ)-DP لزوماً ایجاب نمی‌کند که $\Pr[L\geq\varepsilon]\leq\delta$ باشد. در واقع وجود دارند (ε,δ)-DP مکانیزم‌هایی که برای آنها $\Pr[L\geq\varepsilon]>\delta$ است.

در بخش (ب) ثابت کردیم که $\boxed{\text{tail bound condition}\Rightarrow(\varepsilon,\delta)\text{-DP}}$ ولی برعکس این قضیه در حالت کلی برقرار نیست.

چرا این شرط دقیق (tight) نیست؟ مشخصه سازی واگرایی هاکی-استیک در بخش الف) یک مشخصه سازی دقیق و exact است:

$$H_{e^{\varepsilon}}(M(X')\|M(X))=\int[f_{X'}(y)-e^{\varepsilon}f_X(y)]_+dy\leq\delta$$

ولی tail bound دقیق نیست:

$$\Pr[L_{X'\|X}(Y')\geq\varepsilon]=\int_{\{y:f_{X'}(y)\geq e^{\varepsilon}.f_X(y)\}} f_{X'}(y)dy\leq\delta$$

CamScanner

تفاوت: هاکی - استیک excess mass یا جرم اضافه $\left[f_{X'}(y) - e^{\varepsilon} f_X(y)\right]_+$ را اندازه گیری می کند در حالی که tail bound جرم کل (total mass) $f_{X'}(y)$ در ناحیه‌ای که نسبت چگالی احتمالات بزرگ است اندازه‌گیری می کند. حال یک مثال نقض می زنیم.

مثال نقض: یک مکانیزم را در نظر می گیریم که در ناحیه $\{y : L_{X'\|X}(y) \geq \varepsilon\}$ داریم:

$$f_{X'}(y) = e^{\varepsilon} \cdot f_X(y) + \text{small excess}$$

توجه کنید که tail probability در این مثال می تواند خیلی بزرگ تر از δ باشد ولی واگرایی هاکی - استیک همچنان می تواند کوچکتر مساوی δ باشد. $H_{e^{\varepsilon}}(M(X')\|M(X)) \leq \delta$

بنابراین tail bound یک شرط کافی ولی نه لازم بر (ε,δ)-DP است. توجه کنید tail bound یک تکنیک اثبات مفید است چراکه محاسبه tail probabilities معمولا ساده تر از محاسبه واگرایی هاکی - استیک است ولی مقداری از tightness را از دست می دهد.

مثال نقض محسوس تر: یک مکانیزم را در نظر بگیرید که با احتمال δ مقداری را خروجی می دهد که کاملاً مجموعه داده را آشکار می کند و با احتمال $1-\delta$ خروجی ای می دهد که ε-DP است. مقدار privacy loss روی رویداد بد و نقض کننده محرمانگی تفاضلی ممکن است بزرگ باشد ولی مکانیزم همزمان (ε,δ)-DP است چراکه رویداد بد با احتمال حداکثر δ رخ می دهد. ولی $\Pr[L \geq \varepsilon]$ می تواند به اندازه $\delta + (1-\delta).\Pr[L \geq \varepsilon \text{ in the good part}]$ بزرگ باشد و ممکن است بزرگ تر از δ شود.

سوال ۴: مکانیزم گاوسی: برای اثبات این قضیه از کتاب The Algorithmic Foundation of Differential Privacy از Cynthia Dwork و Aaron Roth کمک می گیریم. در قسمت A از Appendices این اثبات موجود است. توجه کنید این قضیه را می توان به شکل زیر (همان طور که در اسلایدهای درس و کتاب ذکر شده آمده است) بیان کرد که در آن از $\ln(\frac{1.25}{\delta})$ به جای $\log(\frac{1.32}{\delta})$ و ۱ در صورت سوال استفاده شده که تفاوتی ندارند و صورت های مختلف قضیه اند و با تغییر ضرایب و پایه های لگاریتم به هم تبدیل می شوند.

فرض کنید $\varepsilon \in (0,1)$ یک عدد دلخواه باشد. برای $c^2 > 2\ln(\frac{1.25}{\delta})$ مکانیزم گاوسی با پارامتر $\sigma \geq \frac{c.\Delta_2 f}{\varepsilon}$ یک مکانیزم (ε,δ)-DP است. حساسیت ℓ_2 آن هم برابر است با:

$$\Delta_2 f = \max_{\text{adjacent } x,y} \|f(x) - f(y)\|_2 \qquad f : \mathbb{N}^{|X|} \to \mathbb{R}^d$$

CamScanner

یک دیتابیس D وجود دارد به همراه یک پرسش f از آن و مکانیزم گاوسی $f(D)+\eta$ را خروجی می دهد که در آن نویز η از توزیع نرمال می آید. در واقع داریم نویز $N(0,\sigma^2)$ را اضافه می کنیم. در حال حاضر فرض کنید f داریم راجع به توابع real-valued صحبت می کنیم. بنابراین $\Delta f = \Delta_1 f = \Delta_2 f$

ما داریم به این مقدار توجه می کنیم:

$$\left| Ln \frac{e^{(\frac{-1}{2\sigma^2})x^2}}{e^{(\frac{-1}{2\sigma^2})(x+\Delta f)^2}} \right|$$

~~ما در حال بررسی این احتمال به شرط داشتن دیتابیس D و همچنین دیتابیس D'~~

ما در حال بررسی احتمال خروجی هستیم که با احتمال متفاوتی تحت D نسبت به پایگاه داده مجاور D' رخ می دهد، با توجه به اینکه پایگاه D را داریم که در آن فضای احتمال الگوریتم تولید نویز است. صورت کسر در نسبت بالا احتمال مشاهده $f(D)+x$ را در زمانی که پایگاه داده D است توصیف می کند. مخرج کسر مربوط به احتمال مشاهده همین مقدار در زمانی است که پایگاه داده D' است. این نسبت احتمالات است بنابراین همیشه مثبت است اما لگاریتم نسبت ممکن است منفی باشد. متغیر تصادفی مورد نظر ما privacy loss است که عبارت است از

$$Ln \frac{e^{(\frac{-1}{2\sigma^2})x^2}}{e^{(\frac{-1}{2\sigma^2})(x+\Delta f)^2}}$$

و ما به دنبال قدر مطلق آن هستیم:

$$\left| Ln \frac{e^{(\frac{-1}{2\sigma^2})x^2}}{e^{(\frac{-1}{2\sigma^2})(x+\Delta f)^2}} \right| = \left| Ln\, e^{(\frac{-1}{2\sigma^2})(x^2-(x+\Delta f)^2)} \right| = \left| -\frac{1}{2\sigma^2}\left[x^2-(x^2+2x\Delta f+\Delta f^2)\right] \right|$$

$$= \left| \frac{1}{2\sigma^2}\left[2x\Delta f+(\Delta f)^2\right] \right| \quad \text{معادله ①}$$

این کمیت هر زمان که $x < \frac{\sigma^2\varepsilon}{\Delta f} - \frac{\Delta f}{2}$ باشد توسط ε محدود می شود. برای اطمینان از محدود شدن privacy loss توسط ε با احتمال حداقل $1-\delta$ نیاز داریم:

$$Pr\left[|x| \geq \frac{\sigma^2\varepsilon}{\Delta f} - \frac{\Delta f}{2}\right] < \delta$$

و از آن جا که ما با $|x|$ سروکار داریم σ را طوری خواهیم یافت که:

$$Pr\left[x \geq \frac{\sigma^2\varepsilon}{\Delta f} - \frac{\Delta f}{2}\right] < \frac{\delta}{2}$$

ما در کل فرض می کنیم که $\varepsilon \leq 1 \leq \Delta f$ همچنین از tail bound استفاده می کنیم:

$$Pr[x > t] \leq \frac{\sigma}{\sqrt{2\pi}} e^{\frac{-t^2}{2\sigma^2}}$$

CamScanner

ما نیاز داریم که

$$\frac{\sigma}{\sqrt{2\pi}}\frac{1}{t}e^{\frac{-t^2}{2\sigma^2}} < \frac{\delta}{2} \iff \sigma\frac{1}{t}e^{\frac{-t^2}{2\sigma^2}} < \frac{\sqrt{2\pi}\,\delta}{2} \iff \frac{t}{\sigma}e^{\frac{t^2}{2\sigma^2}} > \frac{2}{\sqrt{2\pi}\,\delta}$$

$$\iff \ln\left(\frac{t}{\sigma}\right)+\frac{t^2}{2\sigma^2} > \ln\left(\frac{2}{\sqrt{2\pi}\,\delta}\right)$$

با در نظر گرفتن $t=\frac{\sigma^2\varepsilon}{\Delta f}-\frac{\Delta f}{2}$ خواهیم داشت:

$$\ln\left(\frac{\frac{\sigma^2\varepsilon}{\Delta f}-\frac{\Delta f}{2}}{\sigma}\right)+\frac{\left(\frac{\sigma^2\varepsilon}{\Delta f}-\frac{\Delta f}{2}\right)^2}{2\sigma^2} > \ln\left(\frac{2}{\sqrt{2\pi}\,\delta}\right)=\ln\left(\sqrt{\frac{2}{\pi}}\frac{1}{\delta}\right)$$

اجازه دهید بنویسیم $\sigma=\frac{c\Delta f}{\varepsilon}$ ما می خواهیم c را bound کنیم. با یافتن شرایطی که تحت آنها جمله اول غیرمنفی است، شروع می کنیم.

$$\frac{1}{\sigma}\left(\sigma^2\frac{\varepsilon}{\Delta f}-\frac{\Delta f}{2}\right)=\frac{1}{\sigma}\left[\left(c^2\frac{\Delta f^2}{\varepsilon^2}\right)\frac{\varepsilon}{\Delta f}-\frac{\Delta f}{2}\right]=\frac{1}{\sigma}\left[c^2\left(\frac{\Delta f}{\varepsilon}\right)-\frac{\Delta f}{2}\right]$$

$$=\frac{\varepsilon}{c\Delta f}\left[c^2\left(\frac{\Delta f}{\varepsilon}\right)-\frac{\Delta f}{2}\right]=c-\frac{\varepsilon}{2c}.$$

از آن جایی که $\varepsilon\leq 1$ و $c\geq 1$ مستند داریم $c-\frac{\varepsilon}{2c}\geq c-\frac{1}{2}$ بنابراین $\ln\left(\frac{1}{\sigma}\left(\sigma^2\frac{\varepsilon}{\Delta f}-\frac{\Delta f}{2}\right)\right)>0$ به شرطی که $c\geq\frac{3}{2}$ باشد. بنابراین می توانیم روی عبارت $\frac{t^2}{\sigma^2}$ تمرکز کنیم.

$$\left(\frac{1}{2\sigma^2}\frac{\sigma^2\varepsilon}{\Delta f}-\frac{\Delta f}{2}\right)^2=\frac{1}{2\sigma^2}\left[\Delta f\left(\frac{c^2}{\varepsilon}-\frac{1}{2}\right)\right]^2=\left[(\Delta f)^2\left(\frac{c^2}{\varepsilon}-\frac{1}{2}\right)\right]^2\left[\frac{\varepsilon^2}{c^2(\Delta f)^2}\right]\frac{1}{2}$$

$$=\frac{1}{2}\left(\frac{c^2}{\varepsilon}-\frac{1}{2}\right)^2\frac{\varepsilon^2}{c^2}=\frac{1}{2}\left(\frac{c^2-\varepsilon+\varepsilon^2}{4c^2}\right)$$

چون $\varepsilon\leq 1$ است مشتق $\frac{c^2-\varepsilon+\varepsilon^2}{4c^2}$ نسبت به c در بازه مورد بررسی ما $\left(c\geq\frac{3}{2}\right)$ مثبت است، پس

$\frac{c^2-\varepsilon+\varepsilon^2}{4c^2}\geq c^2-\frac{1}{9}$ و کافی است اطمینان حاصل شود که $c^2-\frac{1}{9}>2\ln\left(\sqrt{\frac{2}{\pi}}\frac{1}{\delta}\right)$

به عبارت دیگر ما به آن نیاز داریم که:

$$c^2>2\ln\left(\sqrt{\frac{2}{\pi}}\right)+2\ln\left(\frac{1}{\delta}\right)+\ln\left(e^{\frac{1}{9}}\right)=\ln\left(\frac{2}{\pi}\right)+\ln\left(e^{\frac{1}{9}}\right)+2\ln\left(\frac{1}{\delta}\right).$$

که از آنجایی که $\left(\frac{2}{\pi}\right)e^{\frac{1}{9}}<1.55$ هرگاه $c^2>2\ln\left(\frac{1.25}{\delta}\right)$ باشد برقرار است.

بگذارید $\mathbb{R}$ را به صورت $\mathbb{R}=R_1\cup R_2$ تقسیم کنیم که در آن $R_1=\left\{x\in\mathbb{R}:|x|\leq\frac{c\Delta f}{\varepsilon}\right\}$, $R_2=\left\{x\in\mathbb{R}:|x|>\frac{c\Delta f}{\varepsilon}\right\}$

هر زیرمجموعه $S\subseteq\mathbb{R}$ را ثابت (fix) می کنیم و تعریف می کنیم:

$S_1=\{f(x)+x \mid x\in R_1\}$ $\quad S_2=\{f(x)+x \mid x\in R_2\}$

پس داریم:

$$\Pr_{x\sim N(0,\sigma^2)}[f(x)+x\in S]=\Pr_{x\sim N(0,\sigma^2)}[f(x)+x\in S_1]+\Pr_{x\sim N(0,\sigma^2)}[f(x)+x\in S_2]$$

$$\le \Pr_{x\sim N(0,\sigma^2)}[f(x)+x\in S_1]+\delta \le e^{\varepsilon}\left(\Pr_{x\sim N(0,\sigma^2)}[f(y)+x\in S_1]\right)+\delta$$

که (ε,δ)-differential privacy را برای مکانیزم گاوسی در یک بعد نتیجه می دهد.

حال برای ابعاد بالا و به شکل برداری تعمیم می دهیم. برای تعمیم این قضیه به توابع $\mathbb{R}^m$ تعریف می کنیم $\Delta f=\Delta_2 f$ اکنون می توانیم این استدلال را با استفاده از norm های اقلیدسی تکرار کنیم.

فرض کنید v هر برداری باشد که $\|v\|\le\Delta f$ را برآورده کند. برای یک جفت ثابت از پایگاه داده های x و y مابه $v=f(x)-f(y)$ علاقه مند هستیم زیرا این چیزی است که نویزما باید آن را پنهان کند.

مانند حالت یک بعدی ما به دنبال شرایطی روی σ هستیم که تحت آن privacy loss زیر

$$\left|\ln \frac{e^{\left(\frac{-1}{2\sigma^2}\right)\|x-\mu\|^2}}{e^{\left(\frac{-1}{2\sigma^2}\right)\|x+v-\mu\|^2}}\right|$$

توسط ε محدود شود. در اینجا x از $N(0,\Sigma)$ انتخاب شده است، که در آن Σ یک ماتریس قطری با درایه های σ^2 است، و از این رو $\mu=(0,\dots,0)$

$\Sigma=\sigma^2 I$

$$\left|\ln \frac{e^{\left(\frac{-1}{2\sigma^2}\right)\|x-\mu\|^2}}{e^{\left(\frac{-1}{2\sigma^2}\right)\|x+v-\mu\|^2}}\right|=\left|\ln e^{\left(\frac{-1}{2\sigma^2}\right)\left[\|x-\mu\|^2-\|x+v-\mu\|^2\right]}\right|=\left|\frac{1}{2\sigma^2}\left(\|x\|^2-\|x+v\|^2\right)\right|$$

ما از این واقعیت استفاده خواهیم کرد که توزیع یک نرمال متقارن کروی (spherically symmetric normal) مستقل از پایه متعامدی است که نرمال های تشکیل دهنده آن از آن گرفته شده اند، پس می توانیم در پایه ای کار کنیم که با v همسو (aligned) باشد. چنین پایه ای را با $b_1,\dots,b_m$ تعیین می کنیم و x را با گرفتن طول های نشان دار $\lambda_i\sim N(0,\sigma^2)$ برای $i\in[m]$ بر می داریم و سپس $x^{[i]}=\lambda_i b_i$ را تعریف می کنیم و در نهایت $x=\sum_{i=1}^{m}x^{[i]}$ را در نظر می گیریم. بدون از دست دادن کلیت فرض می کنیم b_1 موازی v است. ما به $\left|\|x\|^2-\|x+v\|^2\right|$ علاقه مند هستیم.

مثلث قائم الزاویه ای با قاعده $v+x^{[1]}$ و ضلع $\sum_{i=2}^{m}x^{[i]}$ عمود بر v در نظر بگیرید. و تر این مثلث $x+v$ است.

Subject: Year: Month: Date:

$$\|x+v\|^2 = \|v+x^{[1]}\|^2 + \sum_{i=2}^{m} \|x^{[i]}\|^2 \quad , \quad \|x\|^2 = \sum_{i=1}^{m} \|x^{[i]}\|^2$$

از آنجایی که v موازی با $x^{[1]}$ است، داریم: $\|v+x^{[1]}\|^2 = (\|v\| + \lambda_1)^2$

بنابراین

$$\|x+v\|^2 - \|x\|^2 = \|v\|^2 + 2\lambda_1 \cdot \|v\|$$

به یاد بیاورید که $\|v\| \leq \Delta f$ و $\lambda \sim N(0, \sigma^2)$ بنابراین اکنون دقیقاً به حالت

یک بعدی برگشته‌ایم و به‌جای x در معادله (1) λ_1 می‌نویسیم:

$$\left| \frac{1}{2\sigma^2} (\|x\|^2 - \|x+v\|^2) \right| \leq \left| \frac{1}{2\sigma^2} (2\lambda_1 \Delta f - (\Delta f)^2) \right|$$

و بقیه استدلال به شرح فوق ادامه می‌یابد. □