Fixing the DP-GBDT Proof

I. Introduction

Definition 1 (Neighboring databases)**.** *We say that a pair of databases D, D' is neighboring, written as $D \sim_1 D'$, if they differ in at most one element.*

Let $M'(\cdot) = (SVC \circ PS)(\cdot)$ denote the mechanism which first applies a support-vector-based pre-selection $PS(\cdot)$ and afterwards the single-svm-learning algorithm $SVC(\cdot)$ by Chaudhuri et al '11. Let $M(D)$ be the iterative application of $M' = SVC \circ PS$ where in each iteration j the mechanism M' gets as input $D_j := D_{j-1} \setminus PS(D_{j-1}, t_{j-1})$ and $D_0 := D$ and t_{j-1} is the output of M' of the previous round $j-1$. Chaudhuri et al '11 showed that SVC is ε-DP, i.e., for all neighboring D, D' and all trees t:

$$\Pr[SVC(D) = t] \leq \exp(\varepsilon) \Pr[SVC(D') = t].$$

Lemma 2. *Our proposed SVC learning approximation $M(D)$ is $\exp(2\varepsilon)$-DP.*

Proof. We have to show for all neighboring D, D' and all possible SVC ensembles $(t_j)_{j=1}^n$ (arbitrary but fixed):

$$\Pr[M(D) = (t_j)_{j=1}^n] \leq \exp(2\varepsilon) \Pr[M(D') = (t_j)_{j=1}^n] \tag{1}$$

Define $D_0 := D$, $Q_l := PS(D_{l-1}, t_{l-1})$, and $D_l := D_{l-1} \setminus Q_l$. Then, we know that

$$\begin{aligned} &\Pr[M(D) = (t_i)_{j=1}^n] \\ &= \prod_{j=1}^n \Pr[\underbrace{PS(D_{j-1}, t_{j-1}) = Q_j \wedge SVC(Q_j) = t_j}_{=:(t_{j-1}, D_{j-1}, Q_j, t_j)} \mid M_{l=1}^{j-1}(D) = (t_{l=1}^{j-1}, D_{l=1}^{j-1})] \end{aligned} \tag{2}$$

As $(t_1, \ldots, t_n)$ is known to the attacker, it is fixed in the event $M(D) = (t_j)_{j=1}^n$

$$= \prod_{j=1}^n \Pr[(t_{j-1}, D_{j-1}, Q_j, t_j)]. \tag{3}$$

Without loss of generality let $D[s] \neq D'[s]$ and $\forall_{q \neq s} D[q] = D'[q]$. In other words $D \cap D' = D \setminus \{D[s]\} = D' \setminus \{D'[s]\}$. We next prove the statement by induction.

We identify an invariant

$$\bigvee_{q=1}^4 A_q$$

that we show to hold at each iteration $j \in \{1, \ldots, n\}$. The invariant states that one of the following four properties holds at each step.

$$\begin{aligned} A_1 :\Leftrightarrow\ & D'_j \cap D_j = D_j \setminus \{D_j[s]\} = D'_j \setminus \left\{D'_j[s]\right\} \\ & \wedge \prod_{l=1}^j \Pr[M(D) = t_l] = \prod_{l=1}^j \Pr[M(D') = t_l] \end{aligned} \tag{4}$$

$$\begin{aligned} A_2 :\Leftrightarrow\ & D'_j \cap D_j = D_j \setminus \{D_j[s]\} = D'_j \\ & \wedge \prod_{l=1}^j \Pr[M(D) = t_l] \leq \exp(\varepsilon) \prod_{l=1}^j \Pr[M(D') = t_l] \end{aligned} \tag{5}$$

$$\begin{aligned} A_3 :\Leftrightarrow\ & D'_j \cap D_j = D'_j \setminus \left\{D'_j[s]\right\} = D_j \\ & \wedge \prod_{l=1}^j \Pr[M(D) = t_l] \leq \exp(\varepsilon) \prod_{l=1}^j \Pr[M(D') = t_l] \end{aligned} \tag{6}$$

$$\begin{aligned} A_4 :\Leftrightarrow\ & D'_j \cap D_j = D'_j = D_j \\ & \wedge \prod_{l=1}^j \Pr[M(D) = t_l] \leq \exp(2\varepsilon) \prod_{l=1}^j \Pr[M(D') = t_l] \end{aligned} \tag{7}$$

a) Base case:: For $D'_0 := D'$, we have three cases for $Q'_0 := PS(D'_0, \emptyset)$

$$Q'_0 \cap Q_0 = \begin{cases} Q'_0 \\ Q'_0 \setminus \{D'[s]\} \\ Q_0 \setminus \{D[s]\} \end{cases} .$$

First, observe that

$$\Pr[M'(D_0) = t_0] = \Pr[SVC(PS(D_0, \emptyset)) = t_0] = \Pr[SVC(Q_0) = t_0]$$

From Chaudhuri at al., we know that for any neighboring Q_0, Q'_0 pair, we have

$$\Pr[SVC(Q_0) = t_0] \leq \exp(\varepsilon) \Pr[SVC(Q'_0) = t_0]$$

Hence, we can conclude

$$\Pr[M(D) = t_0] \begin{cases} = \Pr[SVC(Q'_0) = t_0], & \text{and } Q'_0 \cap Q_0 = Q'_0 \\ \leq \exp(\varepsilon) \Pr[SVC(Q'_0) = t_0], & \text{and } Q'_0 \cap Q_0 = Q'_0 \setminus \{D'[s]\} \\ \leq \exp(\varepsilon) \Pr[SVC(Q'_0) = t_0], & \text{and } Q'_0 \cap Q_0 = Q_0 \setminus \{D[s]\} \end{cases} .$$

This implies the invariant for $j = 0$.

b) Induction case:: Assume that $\bigvee_{q=1}^{4} A_q$ holds at iteration $i - 1$. We have to show that it then also holds at iteration i. Let us conduct a case distinction over the precondition. For iteration $i - 1$, one of A_1, .., A_4 holds.

$\triangleright$ *If A_1 holds,* we are in the same situation as in the base case. Hence, the induction invariant also holds for step i.

$\triangleright$ *If A_2 holds,* $D_{i-1} \cap D'_{i-1} = D_{i-1} \setminus \{D[s]\} = D'_j$ and

$$\prod_{j=1}^{i-1} \Pr[M(D) = t_j] \leq \exp(\varepsilon) \prod_{j=1}^{i-1} \Pr[M(D') = t_j]$$

We distinguish two subcases here. In the first case, we have

$$PS(D_{i-1}, t_{i-1}) = Q_i = Q'_i = PS(D'_{i-1}, t_{i-1})$$

Then, $\Pr[SVC(Q_i) = t_i] = \Pr[SVC(Q'_i) = t_i]$ and for iteration i the sub-invariant A_2 holds. In the second case, we have

$$PS(D_{i-1}, t_{i-1}) = Q_i = Q'_i \cup \{D[s]\} \text{ and } Q'_i = PS(D'_{i-1}, t_{i-1})$$

Then, as Chaudhuri et al. showed

$$\begin{aligned} \Pr[SVC(Q_i) = t_i] &= \Pr[SVC(Q'_i \cup \{D[s]\}) = t_i] && (8) \\ &\leq \exp(\varepsilon) \Pr[SVC(Q'_i) = t_i] && (9) \\ & && (10) \end{aligned}$$

In this case,

$$\begin{aligned} D_j &= D_{j-1} \setminus Q_j && (11) \\ &= (D'_{j-1} \cup \{D[s]\}) \setminus (Q'_j \cup \{D[s]\}) && (12) \\ &= D'_{j-1} && (13) \\ & && (14) \end{aligned}$$

Moreover,

$$\underbrace{\Pr[(t_{i-1}, D_{i-1}, Q_i, t_i)]}_{\leq \exp(\varepsilon) \Pr[(t_{i-1}, D_{i-1}, Q_i, t_i)]} \underbrace{\prod_{j=1}^{i-1} \Pr[(t_{j-1}, D_{j-1}, Q_j, t_j)]}_{\leq \exp(\varepsilon) \prod_{j=1}^{i-1} \Pr[(t'_{j-1}, D'_{j-1}, Q'_j, t'_j)]} \tag{15}$$

$$\leq \exp(2\varepsilon) \prod_{j=1}^{i} \Pr[(t'_{j-1}, D'_{j-1}, Q'_j, t'_j)] \tag{16}$$

Hence, for iteration i the sub-invariant A_4 holds.

$\triangleright$ *If A_3 holds,* the argumentation is analogous to A_2. If

$$PS(D_{i-1}, t_{i-1}) = Q_i = Q'_i = PS(D'_{i-1}, t_{i-1})$$

holds. then $\Pr[SVC(Q_i) = t_i] = \Pr[SVC(Q'_i) = t_i]$ and for iteration i the sub-invariant A_2 holds. If

$$PS(D_{i-1}, t_{i-1}) = Q_i \text{ and } Q_i \cup \{D'[s]\} = Q'_i = PS(D'_{i-1}, t_{i-1})$$

holds, then for iteration i the sub-invariant A_4 holds.

$\triangleright$ *If A_4 holds,* $D_{i-1} = D'_{i-1}$. Hence,

$$PS(D_{i-1}, t_{i-1}) = Q_i = Q'_i = PS(D'_{i-1}, t_{i-1})$$

Consequently, A_4 holds for iteration i

$$\underbrace{\Pr[(t_{i-1}, D_{i-1}, Q_i, t_i)]}_{=\Pr[(t_{i-1}, D_{i-1}, Q_i, t_i)]} \underbrace{\prod_{j=1}^{i-1} \Pr[(t_{j-1}, D_{j-1}, Q_j, t_j)]}_{\leq \exp(2\varepsilon) \prod_{j=1}^{i-1} \Pr[(t'_{j-1}, D'_{j-1}, Q'_j, t'_j)]} \tag{17}$$

$$\leq \exp(2\varepsilon) \prod_{j=1}^{i} \Pr[(t'_{j-1}, D'_{j-1}, Q'_j, t'_j)] \tag{18}$$

As the induction proof from above shows, from some iteration i onwards, sub-invariant A_4 holds for all $i' \geq i$ holds. As a result, after iteration n the statement of the lemma holds:

$$\Pr[M(D) = (t_j)_{j=1}^n] \leq \exp(2\varepsilon) \Pr[M(D') = (t_j)_{j=1}^n] \tag{19}$$

$\square$