Packing Lower Bounds

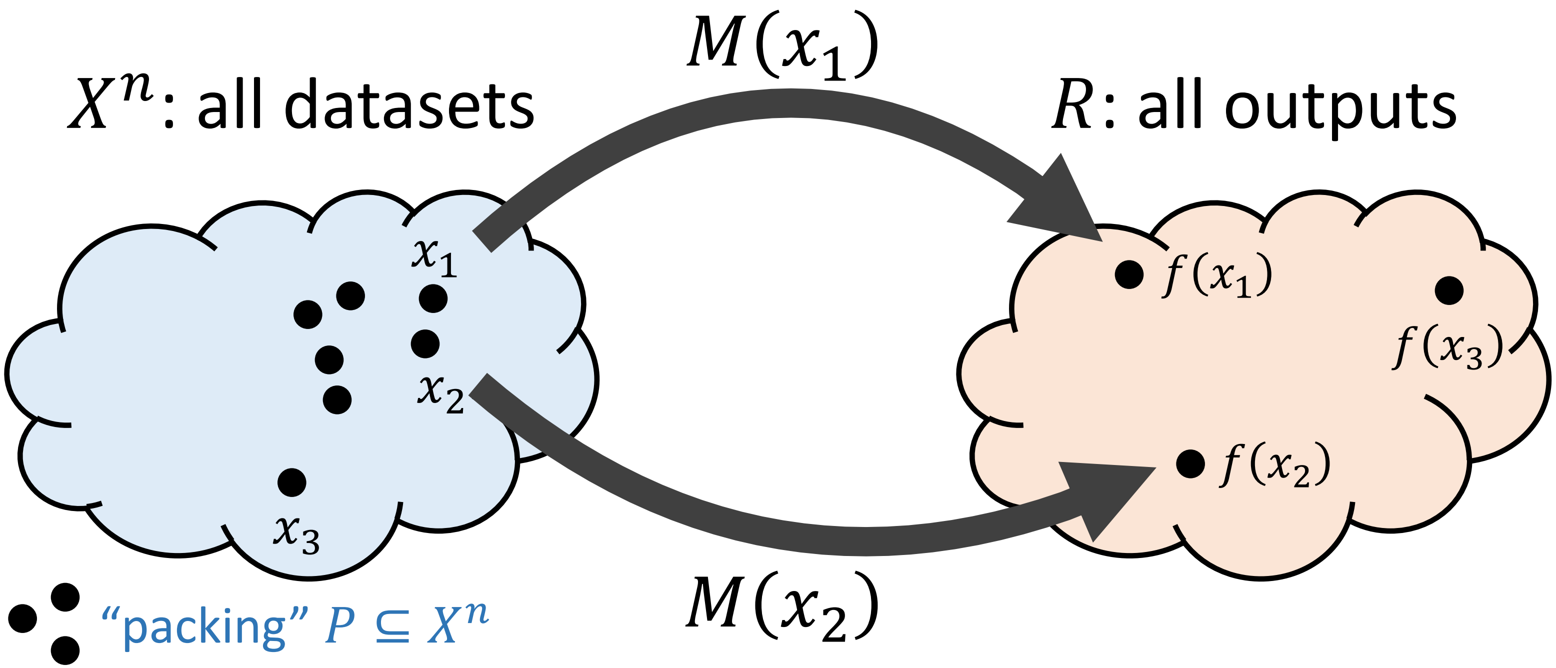
Jonathan Ullman, Northeastern University

Bar Ilan University Winter School: Differential Privacy. February 13, 2017.

Outline

- Packing arguments for DP lower bounds
  - Originated in [HT'10, BKN'10]
  - Intuitive, geometric approach to lower bounds
  - Applicable to a wide variety of problems
  - Often yields tight lower bounds for $(\varepsilon, 0)$-dp
  - Separates $(\varepsilon, 0)$-dp ("pure") from (ε, δ)-dp ("approx")

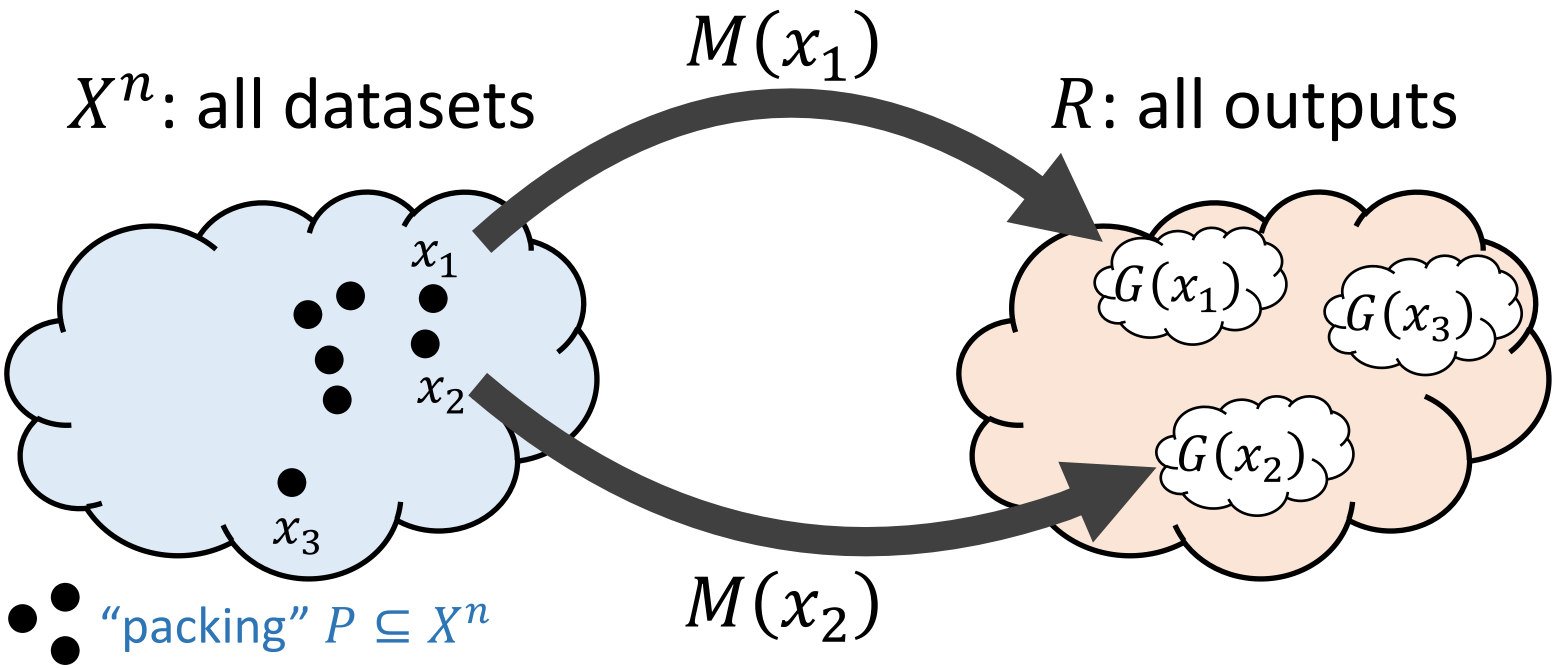
Main Idea



- Find many datasets $P \subseteq X^n$ that are close, but whose answers are far
  - DP implies that $M(x), M(x')$ are close
  - Accuracy implies that $M(x), M(x')$ are far.

Main Idea

$f(x)$ is some function of interest
$G(x)$ are "good outputs for x"



- Find many datasets $P \subseteq X^n$ that are close, but whose answers are far
  - DP implies that $M(x), M(x')$ are close
  - Accuracy implies that $M(x), M(x')$ are far.

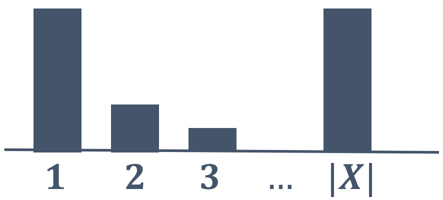
Recall Group Privacy

- Two datasets $x, x' \in X^n$ are neighbors if they differ on at most one row ($x \sim x'$).
- Two datasets $x, x' \in X^n$ are m-neighbors if they differ on at most m rows ($x \sim_m x'$).
- Lemma: If $M : X^n \to R$ is $(\varepsilon, 0)$-differentially private then for every set of m-neighbors $x \sim_m x'$, and every $S \subseteq R$,

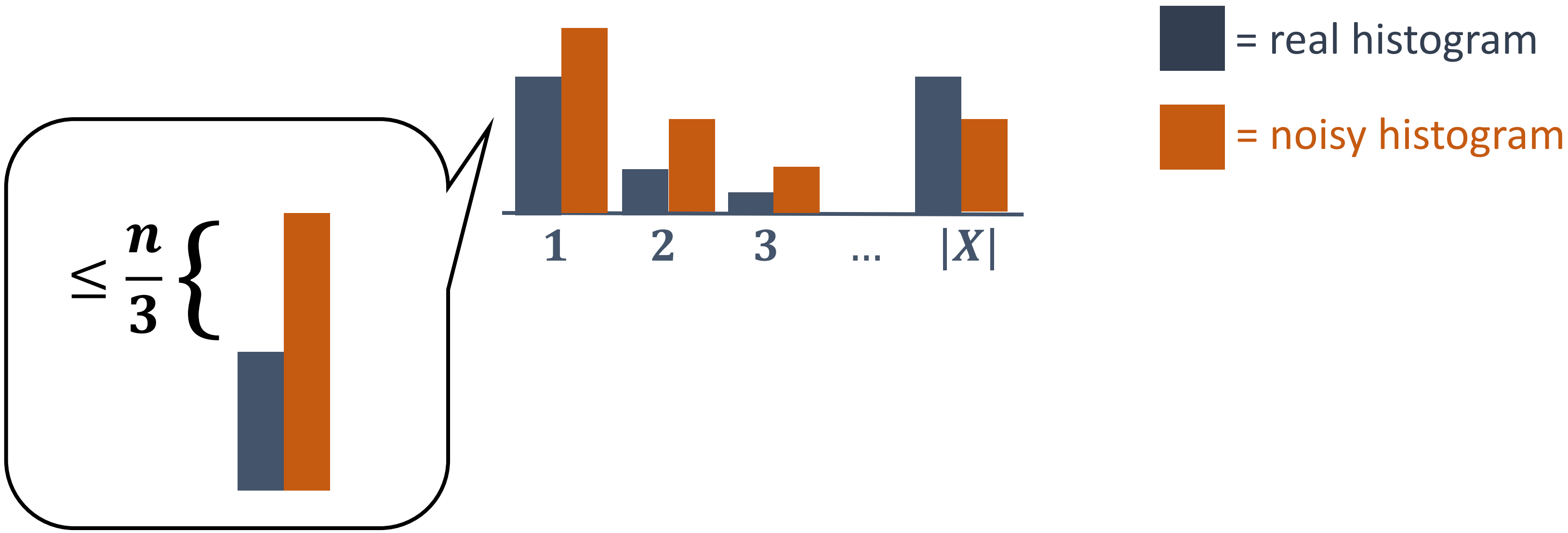
$$\Pr[M(x) \in S] \leq e^{\varepsilon m} \Pr[M(x') \in S]$$

- NB: (ε, δ)-dp doesn't behave as nicely for large groups.

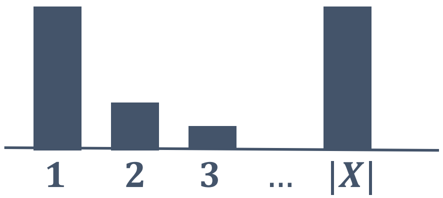
Example: Histograms



- Dataset: $x = (x_1, \dots, x_n) \in X^n$
- Histogram: $h(x)_j = \#\{i : x_i = j\}$
- Accuracy: Release $\hat{h}$ such that $\max_j \left| h(x)_j - \hat{h}_j \right| \le \frac{n}{3}$



Example: Histograms



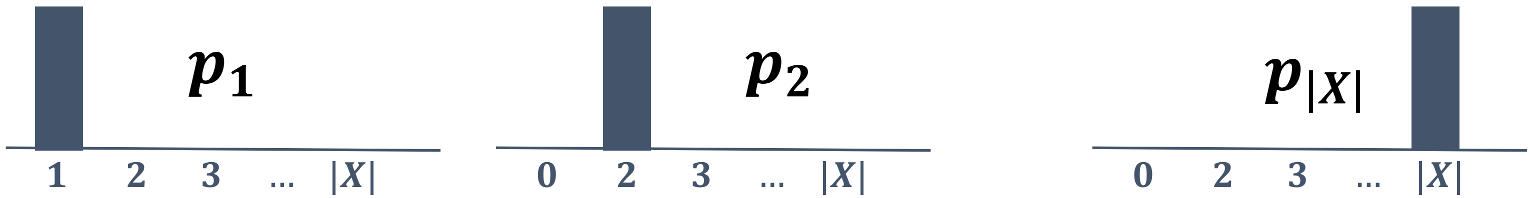
- Dataset: $x = (x_1, \dots, x_n) \in X^n$
- Histogram: $h(x)_j = \#\{i : x_i = j\}$
- Accuracy: $G(x) = \left\{\hat{h} \mid \max_z \left|\hat{h}_z - h(x)_z\right| \leq \frac{n}{3}\right\}$

- Q1: Suppose we use Laplace, how much noise do we need?
- A1: Global ℓ_1-sensitivity is 1, add $\mathrm{Lap}\left(\frac{1}{\varepsilon}\right)$ to each entry

- Q2: How big must n be to satisfy accuracy?
- A2: Largest entry has error $\Theta\left(\frac{\ln|X|}{\varepsilon}\right)$ whp. So $n = \Theta\left(\frac{\ln|X|}{\varepsilon}\right)$ is sufficient for accuracy.

Example: Histograms

- Thm: If $M : X^n \rightarrow \mathbb{N}^{|X|}$ is $(\varepsilon, 0)$-differentially private and $\Pr[M(x) \text{ is an accurate histogram}] \geq \frac{1}{e}$, then $n \geq \frac{\ln|X| - 1}{\varepsilon}$
- Proof: Define the following "packing" of $|X|$ datasets:



- No histogram is "good" for both p and p'
  - $G(p_1), \ldots, G(p_{|X|})$ are mutually disjoint

Example: Histograms

- Thm: If $M : X^n \to \mathbb{N}^{|X|}$ is $(\varepsilon, 0)$-differentially private and $\Pr[M(x) \text{ is an accurate histogram}] \geq \frac{1}{e}$, then $n \geq \frac{\ln|X| - 1}{\varepsilon}$
- Proof:

$$1 \geq \sum_z \Pr[M(p_1) \in G(p_z)]$$ (disjointness)

$$\geq \sum_z e^{-\varepsilon n} \Pr[M(p_1) \in G(p_z)]$$ (group privacy, size n)

$$\geq \sum_z e^{-\varepsilon n} \frac{1}{e}$$ (accuracy)

$$= |X| e^{-\varepsilon n - 1}$$ (size of packing is $|X|$)

$$\Longrightarrow n \geq \frac{\ln|X| - 1}{\varepsilon}$$

General Packing Lemma

- Let $\{G(x)\}_{x \in X^n}$ be a family of subsets of the output range R
  - These are the "good outputs for x"
- m-Packing: Let $P = \{x_0, x_1, \dots\} \subseteq X^n$ be such that
  - every $x, x' \in P$ are m-neighbors (datasets are close)
  - $G(p_0), G(p_1), \dots$ are mutually disjoint (answers are far)

Lemma: If P is an m-packing and $M : X^n \to R$ is an $(\varepsilon, 0)$-dp algorithm such that $\Pr[M(x) \in G(x)] \geq \frac{1}{e}$, then $m \geq \frac{\ln|P| - 1}{\varepsilon}$

Packing Lemma

- Lemma: If $M : X^n \to R$ is $(\varepsilon, 0)$-differentially private, and $\forall x \in P, \Pr[M(x) \in G(x)] \geq \frac{1}{e}$, then $m \geq \frac{\ln|P| - 1}{\varepsilon}$
- Proof:

$$1 \geq \sum_z \Pr[M(x_0) \in G(x_z)] \quad \text{(disjointness)}$$

$$\geq \sum_z e^{-\varepsilon m} \Pr[M(x_z) \in G(x_z)] \quad \text{(group privacy, size } m\text{)}$$

$$\geq \sum_z e^{-\varepsilon m} \frac{1}{e} \quad \text{(accuracy)}$$

$$= |P| e^{-\varepsilon m - 1} \quad \text{(size of packing)}$$

$$\Rightarrow m \geq \frac{\ln|P| - 1}{\varepsilon}$$

Example: Dataset Mean

- Dataset: $x = (x_1, \ldots, x_n) \in \left(\{0,1\}^d\right)^n$
- Mean: $\mu(x) = \frac{1}{n}\sum_i x_i$
- Accuracy: $G(x) = \left\{ \hat{\mu} \;\middle|\; \max_c |\mu(x)_c - \hat{\mu}_c| \leq \alpha \right\}$

- Q1: Suppose we use Laplace, how much noise do we add?
- A1: Global ℓ_1-sensitivity is $\frac{d}{\varepsilon n}$, add $\mathrm{Lap}\left(\frac{d}{\varepsilon n}\right)$ to each entry

- Q2: How accurate?
- A2: $\alpha = O\left(\frac{d \ln d}{\varepsilon n}\right)$ whp. Can be improved to $\alpha = O\left(\frac{d}{\varepsilon n}\right)$.

Example: Dataset Mean

- Dataset: $x = (x_1, \ldots, x_n) \in \left(\{0,1\}^d\right)^n$
- Mean: $\mu(x) = \frac{1}{n}\sum_i x_i$
- Accuracy: $G(x) = \left\{ \hat{\mu} \;\middle|\; \max_c |\mu(x)_c - \hat{\mu}_c| \leq \alpha \right\}$

dataset x	1	1	0	0
	0	1	0	1
	0	1	0	1
$\mu(x)$	0.333	1.000	0.000	0.667
$\hat{\mu}$	0.360	0.980	0.045	0.700

$\alpha = .045$

Example: Dataset Mean

- Dataset: $x = (x_1, \dots, x_n) \in \left(\{0,1\}^d\right)^n$
- Mean: $\mu(x) = \frac{1}{n}\sum_i x_i$
- Accuracy: $G(x) = \left\{ \hat{\mu} \,\middle|\, \max_c |\mu(x)_c - \hat{\mu}_c| \leq \alpha \right\}$
- Define the following packing:
  - $P = \{p_z\}_{z \in \{0,1\}^d}$
  - $|P| = 2^d$
  - p, p' are $m = 3\alpha n$ neighbors
  - $G(x) = \{\hat{\mu} \mid \|\mu(x) - \hat{\mu}\|_\infty \leq \alpha\}$
- Packing lemma $\Rightarrow 3\alpha n \geq \frac{d-1}{\varepsilon}$

Ex: p_{1001}

$3\alpha n$ rows	1001
	1001
	1001
$n - 3\alpha n$ rows	0000
	0000

$\mu(p_z) = 3\alpha z$

Example: Dataset Mean

- Dataset: $x = (x_1, \dots, x_n) \in \left(\{0,1\}^d\right)^n$
- Mean: $\mu(x) = \frac{1}{n}\sum_i x_i$
- Accuracy: $G(x) = \left\{ \hat{\mu} \;\middle|\; \max_c |\mu(x)_c - \hat{\mu}_c| \le \alpha \right\}$
- Define the following packing:
    - $P = \{p_z\}_{z \in \{0,1\}^d}$
    - $|P| = 2^d$
    - p, p' are $m = 3\alpha n$ neighbors
- Packing lemma $\Rightarrow \alpha \ge \frac{d-1}{3\varepsilon n}$

Ex: p_{1001}

$3\alpha n$ rows	1001
	1001
	1001
$n - 3\alpha n$ rows	0000
	0000

$\mu(p_z) = 3\alpha z$

Statistical Queries (SQs)

- Recall statistical queries $q(x) = \frac{1}{n}\sum_i \phi(x_i)$
- The mean is d statistical queries on $x \in \left(\{0,1\}^d\right)^n$

- Thm: Laplace noise is $(\varepsilon, 0)$-dp and answers k arbitrary SQs up to error $\alpha = O\left(\frac{k \ln k}{\varepsilon n}\right)$

- Thm: No $(\varepsilon, 0)$-dp algorithm $M : X^n \to R$ can answer $k \leq \log|X|$ arbitrary SQs with $\alpha < \frac{k-1}{3\varepsilon n}$

Statistical Queries (SQs)

- Recall statistical queries $q(x) = \frac{1}{n}\sum_i \phi(x_i)$
- The mean is d statistical queries on $x \in \left(\{0,1\}^d\right)^n$

- Thm: Laplace noise is (ε, δ)-dp and answers k arbitrary SQs up to error $\alpha = \tilde{O}\left(\frac{\sqrt{k \ln(1/\delta)}}{\varepsilon n}\right)$

- Packing lower bound is false for approximate dp.
- Later on we'll see how to show tight lower bounds for (ε, δ)-dp using very different techniques

Example: Online Counting

- Data: stream of bits $x_1, \ldots, x_T \in \{0,1\}^T$, given one at a time
- Goal: after x_t, output a_t approximating $c_t = \sum_{t' \leq t} x_{t'}$
- Accuracy: $\max_t |a_t - c_t| \leq \alpha$

- Fact: there is an $(\varepsilon, 0)$-dp algorithm with accuracy $\alpha = O(\varepsilon^{-1} \ln T)$. (Binary tree gives $\alpha = O(\varepsilon^{-1} \ln^2 T)$.)

- Theorem: for every $(\varepsilon, 0)$-dp algorithm $\alpha = \Omega(\varepsilon^{-1} \ln T)$.

Example: Online Counting

- Theorem: for every $(\varepsilon, 0)$-dp algorithm $\alpha = \Omega(\varepsilon^{-1} \ln T)$.

Split the input into $B = \frac{T}{3\alpha}$ blocks of length 3α.

$\boldsymbol{p_j}$	**00000**	**11111**	**00000**	**00000**	**00000**

block j

- $P = \left\{ p_j : j = 1, \dots, \frac{T}{3\alpha} \right\}$; $|P| = \frac{T}{3\alpha}$; distance $m = 3\alpha$.
- $G(x) = \left\{ (a_1, \dots, a_T) : \max_t |c_t - a_t| \le \alpha \right\}$
  - $G(p_j), G(p_{j'})$ are disjoint
- Packing Lem. $\Rightarrow m \ge \frac{\ln|P| - 1}{\varepsilon} \Rightarrow 3\alpha \ge \frac{\ln(T) - \ln(3\alpha) - 1}{\varepsilon}$

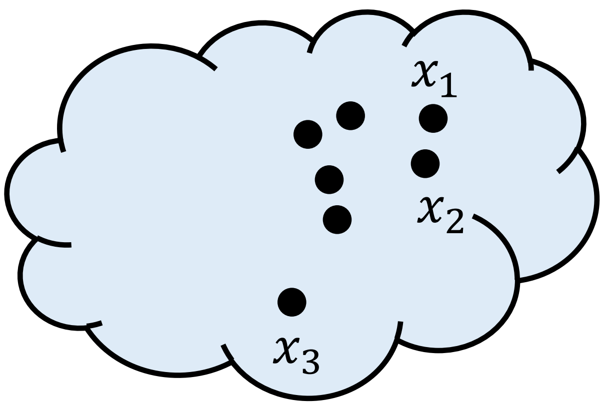
Example: Online Counting

- Theorem: for every $(\varepsilon, 0)$-dp algorithm $\alpha = \Omega(\varepsilon^{-1} \ln T)$.

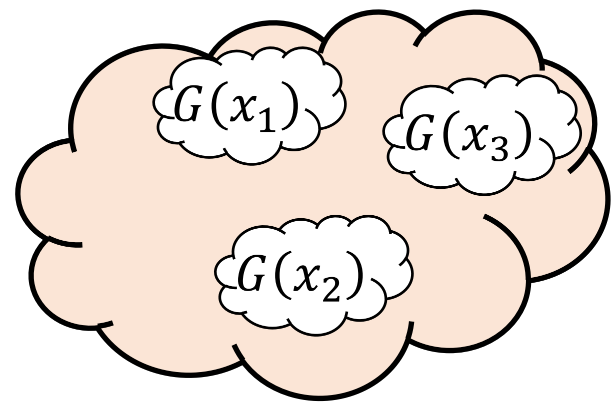
- Also applies to answering threshold queries
  - Dataset $x \in [T]^n$
  - Queries $c_t(x) = \#\{i : x_i \geq t\}$
  - Goal: output $(a_1, \ldots, a_T)$ such that $\max_t |a_t - c_t(x)| \leq \alpha$

Packing vs. Covering

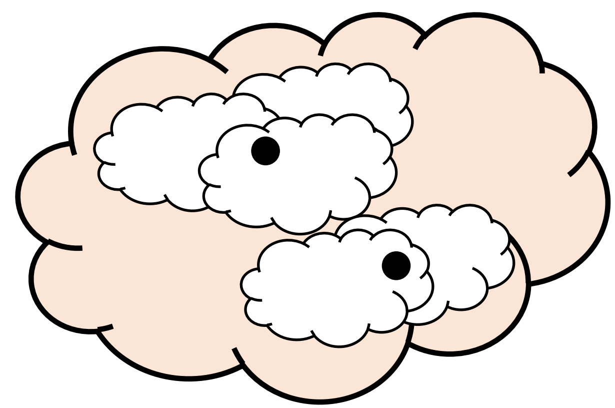
X^n: all datasets



R: all outputs



- packing: set of datasets; no output is “good” for two datasets
- covering: set of outputs; for every dataset, some output is “good”



Final Thought: Packing vs. Covering

- Suppose we have a function $f : X^n \to R$
- Suppose we have a covering C such that for every x, there exists $c \in C$, such that $d(f(x), c) \leq \alpha$.
  - Some accuracy metric d.
- Thm: Exists an $(\varepsilon, 0)$-dp algorithm with error $\beta = \alpha + \frac{\ln|C|}{\varepsilon n}$.
  - If $n = \Omega\left(\frac{\ln|C|}{\varepsilon\alpha}\right)$, then we get error $\beta = O(\alpha)$

- Thm: size of minimum covering ≈ size of maximum packing
  - Implies LB of $n = \Omega\left(\frac{\ln|C|}{\varepsilon}\right)$; tight up to $O\left(\frac{1}{\alpha}\right)$ factor

Outline

- Packing arguments for DP lower bounds
  - Intuitive geometric approach to lower bounds
  - Applicable to a wide variety of problems
  - Often yields tight lower bounds for $(\varepsilon, 0)$-dp
  - Separates $(\varepsilon, 0)$-dp from (ε, δ)-dp