New Nearly-Optimal Coreset for Kernel Density Estimation

Wai Ming Tai

University of Chicago

Kernel Density Estimation

Given:

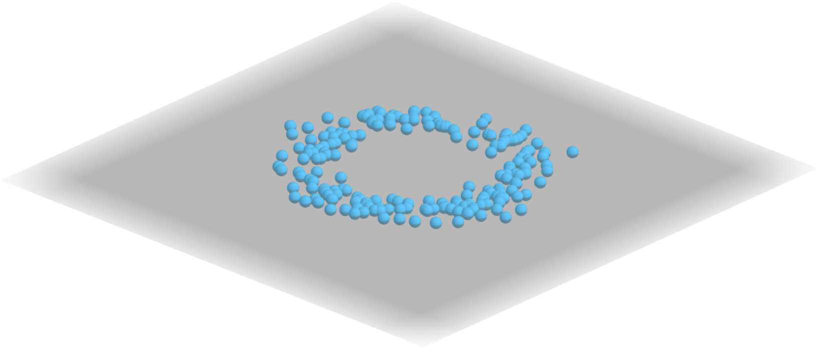
- a set $P \subset \mathbb{R}^d$ of size n
- a kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ (Common example is Gaussian kernel $K(x, y) = e^{-\|x-y\|^2}$ and it is the focus in this talk)
- a query $x \in \mathbb{R}^d$

We want to compute:

$$\overline{\mathcal{G}}_P(x) = \frac{1}{n} \sum_{p \in P} e^{-\|x-p\|^2}$$

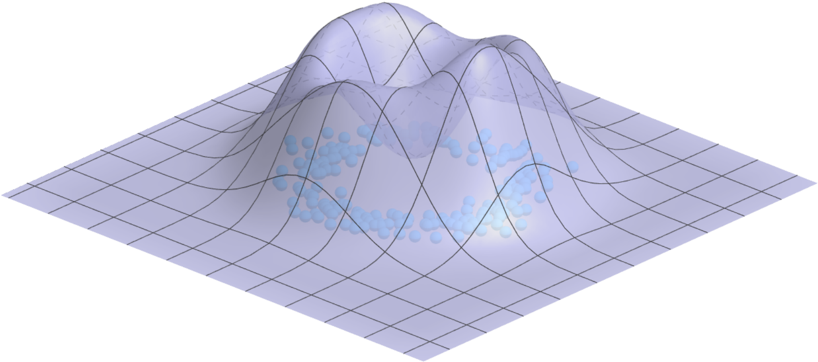
Kernel Density Estimation

Dataset P



Kernel Density Estimation

Kernel Density Estimation of P $\overline{\mathcal{G}}_P(x)$



Kernel Density Estimation

Some Applications:

- Statistics
  - Use kernel density estimation to approximate the unknown distribution [Silverman 1986, Scott 1992]
- Kernel methods in Machine Learning
  - Define distance of two point sets by using kernel density estimation, Kernel SVM, Kernel PCA [Scholkopf+Smola 2002]
- Topological data analysis
  - Study the level set of kernel density estimation and analysis its topological structure [Phillips et al. SOCG 2015]

Problem Definition

The input size is too large. Need to reduce the size.

Problem Definition

The input size is too large. Need to reduce the size.

Find a ε-coreset $Q \subset P$ s.t.

$$\max_{x \in \mathbb{R}^d} \left| \overline{\mathcal{G}}_P(x) - \overline{\mathcal{G}}_Q(x) \right| \leq \varepsilon$$

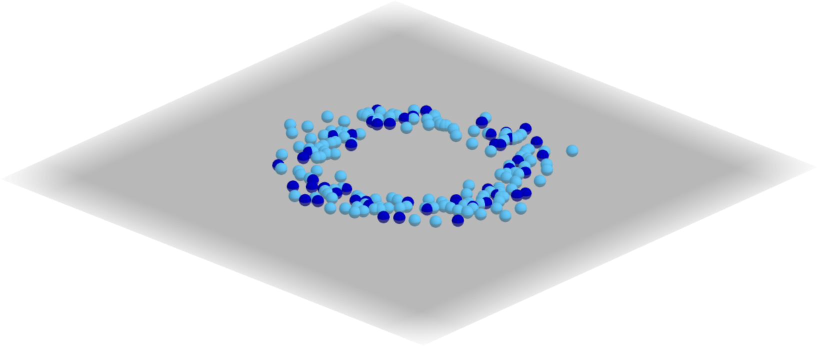
That is,

$$\max_{x \in \mathbb{R}^d} \left| \frac{1}{|P|} \sum_{p \in P} e^{-\|x-p\|^2} - \frac{1}{|Q|} \sum_{q \in Q} e^{-\|x-q\|^2} \right| \leq \varepsilon$$

Question: how small can the size of Q, $|Q|$, be?

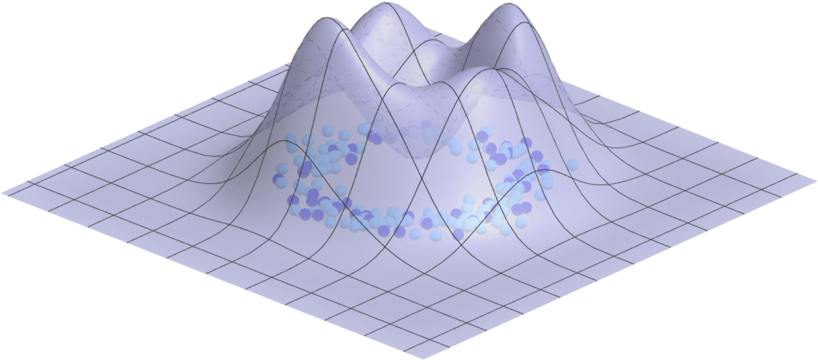
Coreset

Coreset Q



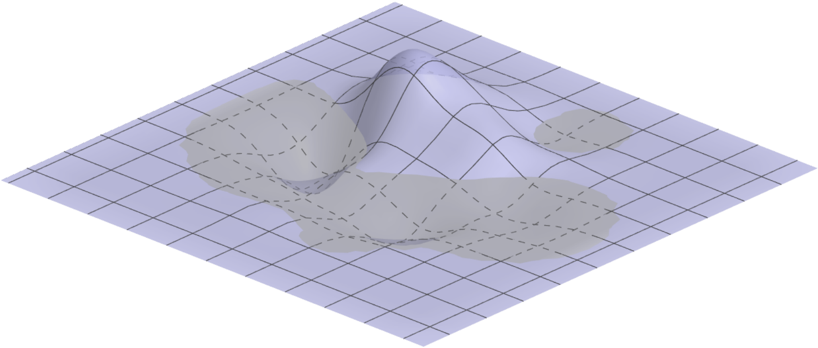
Coreset

Kernel Density Estimation of Q $\overline{\mathcal{G}}_Q(x)$



Coreset

Error $\overline{\mathcal{G}}_P(x) - \overline{\mathcal{G}}_Q(x)$



Previous Results

Highlight of previous results on size of ε-coreset:

Paper	Size	d
Joshi et al. SOCG 2011	$O(d/\varepsilon^2)$	any
Bach et al. ICML 2012	$O(1/\varepsilon^2)$	any
Joshi et al. SOCG 2011	$O(1/\varepsilon)$	1
Phillips SODA 2013	sub-$O(1/\varepsilon^2)$	constant
Phillips+Tai SODA 2018	$O(\frac{1}{\varepsilon}\log^d \frac{1}{\varepsilon})$	constant
Phillips+Tai SOCG 2018	$O(\frac{1}{\varepsilon}\sqrt{d \log \frac{1}{\varepsilon}})$	any
Phillips SODA 2013	$\Omega(1/\varepsilon)$	1
Phillips+Tai SODA 2018	$\Omega(1/\varepsilon^2)$	$\leq 1/\varepsilon^2$
Phillips+Tai SOCG 2018	$\Omega(\sqrt{d}/\varepsilon)$	$> 1/\varepsilon^2$

Previous Results

Highlight of previous results on size of ε-coreset:

Paper	Size	d
Joshi et al. SOCG 2011	$O(d/\varepsilon^2)$	any
Bach et al. ICML 2012	$O(1/\varepsilon^2)$	any
Joshi et al. SOCG 2011	$O(1/\varepsilon)$	1
Phillips SODA 2013	sub-$O(1/\varepsilon^2)$	constant
Phillips+Tai SODA 2018	$O(\frac{1}{\varepsilon}\log^d \frac{1}{\varepsilon})$	constant
Phillips+Tai SOCG 2018	$O(\frac{1}{\varepsilon}\sqrt{d \log \frac{1}{\varepsilon}})$	any
Phillips SODA 2013	$\Omega(1/\varepsilon)$	1
Phillips+Tai SODA 2018	$\Omega(1/\varepsilon^2)$	$\leq 1/\varepsilon^2$
Phillips+Tai SOCG 2018	$\Omega(\sqrt{d}/\varepsilon)$	$> 1/\varepsilon^2$

Our Result: $O(\frac{1}{\varepsilon}\sqrt{\log^* \frac{1}{\varepsilon} \log\log^* \frac{1}{\varepsilon}})$ when d is a constant

Equivalent Problem

By standard halving technique, the following problem is equivalent to our problem definition.

Given:

- Point set $P \in \mathbb{R}^d$

Find:

- Coloring $\sigma : P \to \{-1, +1\}$

Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

Equivalent Problem

Equivalence:

- Our problem: fix the error , minimize the size
- Equivalent problem: fix the size , minimize the error

Equivalent Problem

Equivalence:

- Our problem: fix the error , minimize the size
- Equivalent problem: fix the size , minimize the error

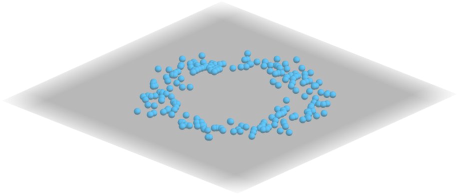
$$\frac{1}{|P|}\left|\sum_{p\in P}\sigma(p)e^{-\|x-p\|^2}\right| = \left|\frac{1}{|P|}\sum_{p\in P}e^{-\|x-p\|^2} - \frac{1}{|P|/2}\sum_{p\in P_+}e^{-\|x-p\|^2}\right|$$

Equivalent Problem

Equivalence:

- Our problem: fix the error , minimize the size
- Equivalent problem: fix the size , minimize the error

$$\frac{1}{|P|}\left|\sum_{p\in P}\sigma(p)e^{-\|x-p\|^2}\right| = \left|\frac{1}{|P|}\sum_{p\in P}e^{-\|x-p\|^2} - \frac{1}{|P|/2}\sum_{p\in P_+}e^{-\|x-p\|^2}\right|$$

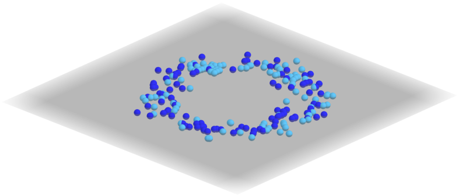


Equivalent Problem

Equivalence:

- Our problem: fix the error , minimize the size
- Equivalent problem: fix the size , minimize the error

$$\frac{1}{|P|}\left|\sum_{p\in P}\sigma(p)e^{-\|x-p\|^2}\right| = \left|\frac{1}{|P|}\sum_{p\in P}e^{-\|x-p\|^2} - \frac{1}{|P|/2}\sum_{p\in P_+}e^{-\|x-p\|^2}\right|$$

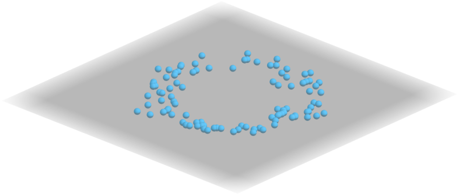


Equivalent Problem

Equivalence:

- Our problem: fix the error , minimize the size
- Equivalent problem: fix the size , minimize the error

$$\frac{1}{|P|}\left|\sum_{p\in P}\sigma(p)e^{-\|x-p\|^2}\right| = \left|\frac{1}{|P|}\sum_{p\in P}e^{-\|x-p\|^2} - \frac{1}{|P|/2}\sum_{p\in P_+}e^{-\|x-p\|^2}\right|$$

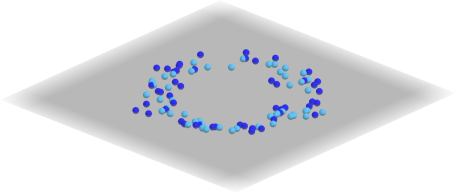


Equivalent Problem

Equivalence:

- Our problem: fix the error , minimize the size
- Equivalent problem: fix the size , minimize the error

$$\frac{1}{|P|}\left|\sum_{p\in P}\sigma(p)e^{-\|x-p\|^2}\right| = \left|\frac{1}{|P|}\sum_{p\in P}e^{-\|x-p\|^2} - \frac{1}{|P|/2}\sum_{p\in P_+}e^{-\|x-p\|^2}\right|$$

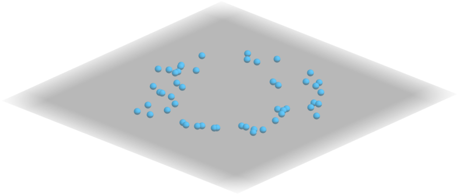


Equivalent Problem

Equivalence:

- Our problem: fix the error , minimize the size
- Equivalent problem: fix the size , minimize the error

$$\frac{1}{|P|}\left|\sum_{p\in P}\sigma(p)e^{-\|x-p\|^2}\right| = \left|\frac{1}{|P|}\sum_{p\in P}e^{-\|x-p\|^2} - \frac{1}{|P|/2}\sum_{p\in P_+}e^{-\|x-p\|^2}\right|$$



Problem Definition

Now, our problem definition becomes
Given:

- Point set $P \in \mathbb{R}^d$

Find:

- Coloring $\sigma : P \to \{-1, +1\}$

Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

Problem Definition

Now, our problem definition becomes
Given:

- Point set $P \in \mathbb{R}^d$

Find:

- Coloring $\sigma : P \to \{-1, +1\}$

Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

Our target size of coreset is

- $O(\frac{1}{\varepsilon}\sqrt{\log^* \frac{1}{\varepsilon} \log\log^* \frac{1}{\varepsilon}})$

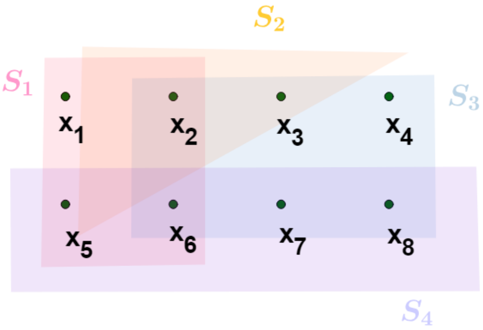
Our goal becomes

- $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right| = O(\sqrt{\log^* n \log\log^* n})$

Discrepancy Theory

Given a set system:

- $X = \{x_i \mid i = 1, 2, \ldots, n\}$
- $\mathcal{S} = \{S_i \mid i = 1, 2, \ldots, m\}$ where S_i is subset of X



Discrepancy Theory

For a given coloring $\sigma : \{1, 2, \ldots, n\} \to \{-1, +1\}$, define the discrepancy of a set S

$$\mathrm{disc}(\sigma, S) = \left| \sum_{x_i \in S} \sigma(x_i) \right|$$

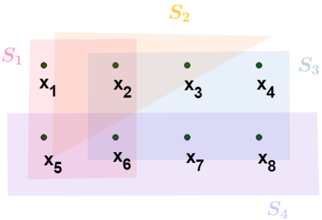
and the discrepancy of a set system $(X, \mathcal{S})$

$$\mathrm{disc}(\mathcal{S}) = \min_{\sigma} \max_{S \in \mathcal{S}} \mathrm{disc}(\sigma, S)$$

Goal: minimize $\mathrm{disc}(\mathcal{S})$

Matrix Representation

Suppose A is a m-by-n matrix s.t. $A_{i,j} = \begin{cases} 1 & \text{if } x_j \in S_i \\ 0 & \text{otherwise.} \end{cases}$



$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Define discrepancy of A

$$\text{disc}(A) = \min_{x \in \{-1,+1\}^n} \|Ax\|_\infty$$

Note: $\text{disc}(\mathcal{S}) = \text{disc}(A)$
Can be generalized to arbitrary matrix

Our Problem

Recall that our problem definition is
Given:

- Point set $P \in \mathbb{R}^d$

Find:

- Coloring $\sigma : P \to \{-1, +1\}$

Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

Our Problem in "matrix" form

Our objective is:

$$\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right| = \|K\sigma\|_\infty$$

where

$$K = {}_{x \in \mathbb{R}^d} \left\{ \overbrace{\begin{bmatrix} & \vdots & \\ \cdots & e^{-\|x-p\|^2} & \cdots \\ & \vdots & \end{bmatrix}}^{p \in P} \right.$$

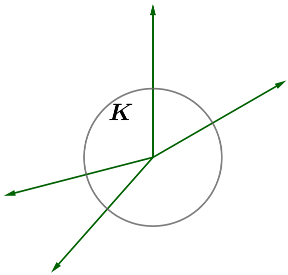
Banaszczyk's Theorem

Banaszczyk's Theorem [Banaszczyk 1998]:
Given:

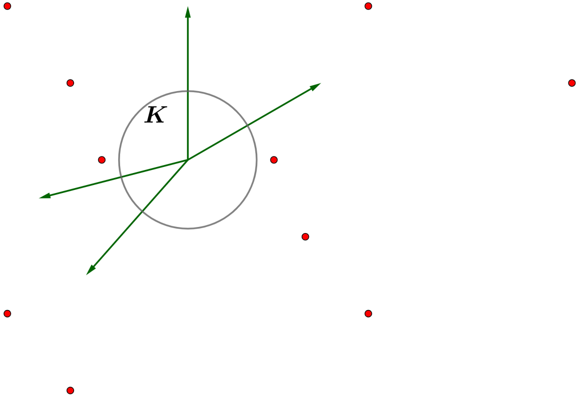
- a convex body $K \in \mathbb{R}^m$ with Gaussian measure $\gamma_m(K) > \frac{1}{2}$
- n vectors $v^{(1)}, v^{(2)}, \ldots, v^{(n)} \in \mathbb{R}^m$ such that $\left\|v^{(i)}\right\| \leq 1$

Then: there is a coloring $\sigma : \{1, 2, \ldots, n\} \to \{-1, +1\}$ such that $\sum_{i=1}^{n} \sigma(i) v^{(i)} \in cK$ for some absolute constant c

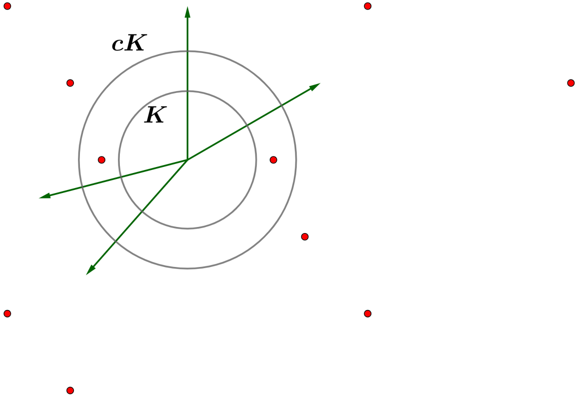
Banaszvzyk's Theorem



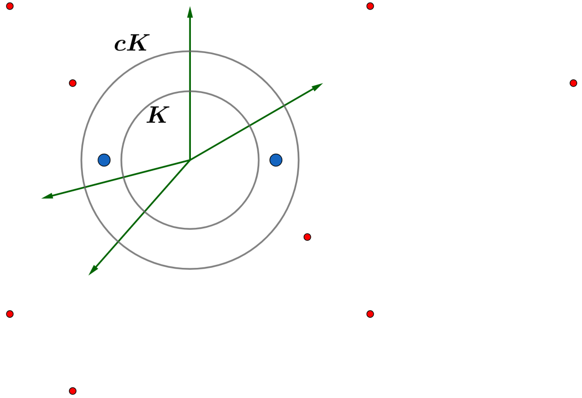
Banaszvzyk's Theorem



Banaszvzyk's Theorem



Banaszvzyk's Theorem



Equivalent statement of Banaszczyk's Theorem

Equivalent statement of Banaszczyk's Theorem [Dadush et al. RANDOM 2016]:

Given:

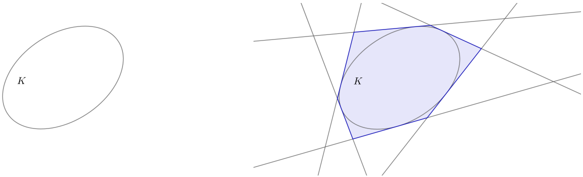
- n vectors $v^{(1)}, v^{(2)}, \ldots, v^{(n)} \in \mathbb{R}^m$ such that $\left\| v^{(i)} \right\| \leq 1$

Then: there is a probability distribution on $\{\sigma : \{1, 2, \ldots, n\} \to \{\pm 1\}\}$ such that there are two absolute constant C_1, C_2 such that, for any unit vector $\theta \in \mathbb{R}^d$ and $\alpha > 0$,

$$\mathbf{Pr}\left[\left| \left\langle \sum_{i=1}^{n} \sigma(i) v^{(i)}, \theta \right\rangle \right| > \alpha \right] < C_1 e^{-C_2 \alpha^2}$$

Why equivalent?

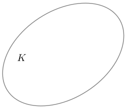
Equivalence:

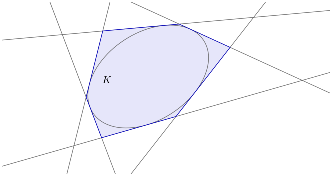


- can always "approximate" a convex body by the intersection of a number of half-spaces
- bound the number of event in union bound

Why equivalent?

Equivalence:





- can always "approximate" a convex body by the intersection of a number of half-spaces
- bound the number of event in union bound

A recent result [Bansal et al. STOC 2018] showed that it is constructive

Our Problem

Our problem definition:

Given: Point set $P \in \mathbb{R}^d$

Find: Coloring $\sigma : P \to \{-1, +1\}$

Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

Our Problem

Our problem definition:
Given: Point set $P \in \mathbb{R}^d$
Find: Coloring $\sigma : P \to \{-1, +1\}$
Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

We want to use Banaszczyk's Theorem:
Given: n vectors $v^{(1)}, v^{(2)}, \ldots, v^{(n)} \in \mathbb{R}^m$ such that $\left\| v^{(i)} \right\| \leq 1$
Then: there is a randomized algorithm to construct a coloring σ such that, for any unit vector $\theta \in \mathbb{R}^d$ and $\alpha > 0$,

$$\mathbf{Pr}\left[\left| \left\langle \sum_{i=1}^{n} \sigma(i) v^{(i)}, \theta \right\rangle \right| > \alpha \right] < O(e^{-\Omega(\alpha^2)})$$

Our Problem

Our problem definition:
Given: Point set $P \in \mathbb{R}^d$
Find: Coloring $\sigma : P \to \{-1, +1\}$
Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

We want to use Banaszczyk's Theorem:
Given: n vectors $v^{(1)}, v^{(2)}, \ldots, v^{(n)} \in \mathbb{R}^m$ such that $\left\| v^{(i)} \right\| \leq 1$
Then: there is a randomized algorithm to construct a coloring σ such that, for any unit vector $\theta \in \mathbb{R}^d$ and $\alpha > 0$,

$$\mathbf{Pr}\left[\left| \left\langle \sum_{i=1}^{n} \sigma(i) v^{(i)}, \theta \right\rangle \right| > \alpha \right] < O(e^{-\Omega(\alpha^2)})$$

But, what is our input vectors $v^{(1)}, v^{(2)}, \ldots, v^{(n)}$?

Positive-Definite Kernel

A symmetric function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called positive-definite kernel if:

- for any $x^{(1)}, x^{(2)}, \ldots, x^{(n)} \in \mathbb{R}^d$ and $c_1, c_2, \ldots, c_n \in \mathbb{R}$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j K(x^{(i)}, x^{(j)}) > 0$$

- Namely, matrix G whose (i, j)-entry is $K(x^{(i)}, x^{(j)})$ is a positive-definite matrix
- G can be decomposed into the form $H^T H$ for some matrix H

Gaussian Kernel is a positive-definite kernel

Our Problem

Our problem definition:

Given: Point set $P \in \mathbb{R}^d$

Find: Coloring $\sigma : P \to \{-1, +1\}$

Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

Our Problem

Our problem definition:

Given: Point set $P \in \mathbb{R}^d$

Find: Coloring $\sigma : P \to \{-1, +1\}$

Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

In "matrix" form:

Our objective is:

$$\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right| = \|K\sigma\|_\infty$$

where

$$K = {}_{x \in \mathbb{R}^d} \left\{ \overbrace{\begin{bmatrix} & \vdots & \\ \cdots & e^{-\|x-p\|^2} & \cdots \\ & \vdots & \end{bmatrix}}^{p \in P} \right.$$

Input for Banaszczyk's Theorem

Important fact: Gaussian Kernel is a positive-definite kernel

Input for Banaszczyk's Theorem

Important fact: Gaussian Kernel is a positive-definite kernel K can be decomposed into the following form.

$$K = \underset{x \in \mathbb{R}^d}{\left\{ \begin{bmatrix} & \vdots & \\ - & u^{(x)} & - \\ & \vdots & \end{bmatrix} \right.} \overbrace{\begin{bmatrix} & | & \\ \cdots & u^{(p)} & \cdots \\ & | & \end{bmatrix}}^{p \in P}$$

Input for Banaszczyk's Theorem

Important fact: Gaussian Kernel is a positive-definite kernel K can be decomposed into the following form.

$$K = {}_{x \in \mathbb{R}^d} \left\{ \begin{bmatrix} & \vdots & \\ - & u^{(x)} & - \\ & \vdots & \end{bmatrix} \right. \overbrace{\begin{bmatrix} & | & \\ \cdots & u^{(p)} & \cdots \\ & | & \end{bmatrix}}^{p \in P}$$

We take $u^{(p)}$ as the input of Banaszczyk's Theorem

Note:

- $u^{(p)}$ has norm 1
- $\left\langle u^{(x)}, u^{(p)} \right\rangle = e^{-\|x-p\|^2}$

Input for Banaszczyk's Theorem

Recall that Banaszczyk's Theorem stated the following:
Given: n vectors $v^{(1)}, v^{(2)}, \ldots, v^{(n)} \in \mathbb{R}^m$ such that $\left\| v^{(i)} \right\| \leq 1$
Then: there is a randomized algorithm to construct a coloring σ such that, for any unit vector $\theta \in \mathbb{R}^d$ and $\alpha > 0$,

$$\mathbf{Pr}\left[\left| \left\langle \sum_{i=1}^{n} \sigma(i) v^{(i)}, \theta \right\rangle \right| > \alpha \right] < O(e^{-\Omega(\alpha^2)})$$

Input for Banaszczyk's Theorem

Recall that Banaszczyk's Theorem stated the following:
Given: n vectors $v^{(1)}, v^{(2)}, \ldots, v^{(n)} \in \mathbb{R}^m$ such that $\left\|v^{(i)}\right\| \leq 1$
Then: there is a randomized algorithm to construct a coloring σ such that, for any unit vector $\theta \in \mathbb{R}^d$ and $\alpha > 0$,

$$\mathbf{Pr}\left[\left|\left\langle \sum_{i=1}^{n} \sigma(i) v^{(i)}, \theta \right\rangle\right| > \alpha\right] < O(e^{-\Omega(\alpha^2)})$$

By Banaszczyk's Theorem, we can construct a coloring σ such that

$$\mathbf{Pr}\left[\left|\left\langle \sum_{p \in P} \sigma(p) u^{(p)}, \theta \right\rangle\right| > \alpha\right] < O(e^{-\Omega(\alpha^2)})$$

Input for Banaszczyk's Theorem

We have:

$$\mathbf{Pr}\left[\left|\left\langle \sum_{p \in P} \sigma(p) u^{(p)}, \theta \right\rangle\right| > \alpha\right] < O\big(e^{-\Omega(\alpha^2)}\big)$$

Input for Banaszczyk's Theorem

We have:

$$\mathbf{Pr}\left[\left|\left\langle \sum_{p \in P} \sigma(p) u^{(p)}, \theta \right\rangle\right| > \alpha\right] < O(e^{-\Omega(\alpha^2)})$$

If we set $\theta = u^{(x)}$, we have

$$\begin{aligned}
&\mathbf{Pr}\left[\left|\sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}\right| > \alpha\right] \\
&= \mathbf{Pr}\left[\left|\left\langle \sum_{p \in P} \sigma(p) u^{(p)}, u^{(x)} \right\rangle\right| > \alpha\right] \\
&< O(e^{-\Omega(\alpha^2)})
\end{aligned}$$

Our Problem

Our problem definition:
Given: Point set $P \in \mathbb{R}^d$
Find: Coloring $\sigma : P \to \{-1, +1\}$
Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$

Our Problem

Our problem definition:
Given: Point set $P \in \mathbb{R}^d$
Find: Coloring $\sigma : P \to \{-1, +1\}$
Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$
From Banaszczyk's Theorem, we have

$$\mathbf{Pr}\left[\left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right| > \alpha \right] < O(e^{-\Omega(\alpha^2)})$$

for one $x \in \mathbb{R}^d$

Our Problem

Our problem definition:
Given: Point set $P \in \mathbb{R}^d$
Find: Coloring $\sigma : P \to \{-1, +1\}$
Goal: minimize $\max_{x \in \mathbb{R}^d} \left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right|$
From Banaszczyk's Theorem, we have

$$\mathbf{Pr}\left[\left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right| > \alpha \right] < O(e^{-\Omega(\alpha^2)})$$

for one $x \in \mathbb{R}^d$
There are infinitely many $x \in \mathbb{R}^d$

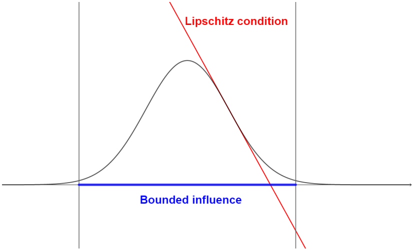
Observations

Lipschitz condition:

- Slope of the kernel cannot be too large

Bounded influence:

- Kernel value is negligible if the query is far away from data



Observations

Assumption: P lies inside a ℓ_∞-ball of radius 1 (we can remove it later)

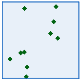
Observations

Assumption: P lies inside a ℓ_∞-ball of radius 1 (we can remove it later)

From the observations:

Observations

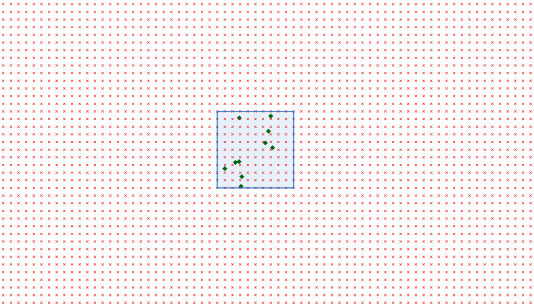
Assumption: P lies inside a ℓ_∞-ball of radius 1 (we can remove it later)
From the observations:



Observations

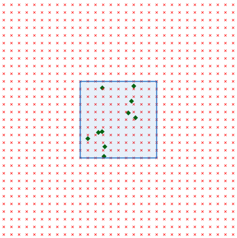
Assumption: P lies inside a ℓ_∞-ball of radius 1 (we can remove it later)

From the observations:



Observations

Assumption: P lies inside a ℓ_∞-ball of radius 1 (we can remove it later)
From the observations:



Observations

How large is the size of this grid?

Observations

How large is the size of this grid?

Denote $D_{\sigma,P}(x) = \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}$

Observations

How large is the size of this grid?

Denote $D_{\sigma,P}(x) = \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}$

From the observations:

- The slope of $D_{\sigma,P}(x)$ is $O(n)$
- $|D_{\sigma,P}(x)| = O(1)$ when $\|x\|_\infty = \Omega(\sqrt{\log n})$

Observations

From the observations:

- The slope of $D_{\sigma,P}(x)$ is $O(n)$
- $|D_{\sigma,P}(x)| = O(1)$ when $\|x\|_\infty = \Omega(\sqrt{\log n})$

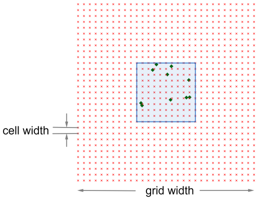
Observations

From the observations:

- The slope of $D_{\sigma,P}(x)$ is $O(n)$
- $|D_{\sigma,P}(x)| = O(1)$ when $\|x\|_\infty = \Omega(\sqrt{\log n})$

It implies:

- The cell width is $\Omega(1/n)$
- The grid width is $O(\sqrt{\log n})$



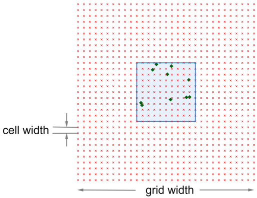
Observations

From the observations:

- The slope of $D_{\sigma,P}(x)$ is $O(n)$
- $|D_{\sigma,P}(x)| = O(1)$ when $\|x\|_\infty = \Omega(\sqrt{\log n})$

It implies:

- The cell width is $\Omega(1/n)$
- The grid width is $O(\sqrt{\log n})$



Size of the grid: $O(n^{O(d)})$

First Attempt

We have:

- $\mathbf{Pr}[|D_{\sigma,P}(x)| > \alpha] < O(e^{-\Omega(\alpha^2)})$ for one $x \in \mathbb{R}^d$
- $|D_{\sigma,P}(x)|$ is small on x in a finite grid of size $O(n^{O(d)})$ implies $|D_{\sigma,P}(x)|$ is small on $x \in \mathbb{R}^d$

First Attempt

We have:

- $\mathbf{Pr}[|D_{\sigma,P}(x)| > \alpha] < O(e^{-\Omega(\alpha^2)})$ for one $x \in \mathbb{R}^d$
- $|D_{\sigma,P}(x)|$ is small on x in a finite grid of size $O(n^{O(d)})$ implies $|D_{\sigma,P}(x)|$ is small on $x \in \mathbb{R}^d$

If we set $\alpha = O(\sqrt{\log n})$ then we have

$$\left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right| = O(\sqrt{\log n})$$

for all $x \in \mathbb{R}^d$

Slightly Strong Result

Recall that $D_{\sigma,P}(x) = \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}$

Slightly Strong Result

Recall that $D_{\sigma,P}(x) = \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}$

It can be rewritten as

- $D_{\sigma,P}(x) = e^{-\frac{2}{3}\|x\|^2} \sum_{p \in P} \sigma(p) e^{2\|p\|^2} e^{-\frac{1}{3}\|x-3p\|^2}$

Slightly Strong Result

Recall that $D_{\sigma,P}(x) = \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}$

It can be rewritten as

- $D_{\sigma,P}(x) = e^{-\frac{2}{3}\|x\|^2} \sum_{p \in P} \sigma(p) e^{2\|p\|^2} e^{-\frac{1}{3}\|x-3p\|^2}$

By the same technique, we have

$$\left| \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2} \right| = O(\sqrt{\log n} \cdot e^{-\Omega(\|x\|^2)})$$

for all $x \in \mathbb{R}^d$

However

$D_{\sigma,P}(x) = \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}$

The reason why we set $\alpha = \sqrt{\log n}$ is

- The slope of $D_{\sigma,P}(x)$ is $O(n)$
- $|D_{\sigma,P}(x)| = O(1)$ when $\|x\|_\infty = \Omega(\sqrt{\log n})$

However

$D_{\sigma,P}(x) = \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}$

The reason why we set $\alpha = \sqrt{\log n}$ is

- The slope of $D_{\sigma,P}(x)$ is $O(n)$
- $|D_{\sigma,P}(x)| = O(1)$ when $\|x\|_\infty = \Omega(\sqrt{\log n})$

Can we do better?

More Observations

$$D_{\sigma,P}(x) = \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}$$

More Observations

$D_{\sigma,P}(x) = \sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}$

For $x, y \in \mathbb{R}^d$, we can express it as

$$|D_{\sigma,P}(x) - D_{\sigma,P}(y)|$$

$$\leq \left| \|y\|^2 - \|x\|^2 \right| |D_{\sigma,P}(x)| + 2 \sum_{j=1}^{d} |y_j - x_j| \left| D_{\sigma,P}(\xi^{(j)}) \right|$$

for some $\xi^{(j)}$ in between x and y

More Observations

We have:

- $|D_{\sigma,P}(x)| = O(\sqrt{\log n} \cdot e^{-\Omega(\|x\|^2)}) = O(\sqrt{\log n})$ for all $x \in \mathbb{R}^d$
- For $x, y \in \mathbb{R}^d$,

$$
\begin{aligned}
&|D_{\sigma,P}(x) - D_{\sigma,P}(y)| \\
&\le \left| \|y\|^2 - \|x\|^2 \right| |D_{\sigma,P}(x)| + 2\sum_{j=1}^{d} |y_j - x_j| \left| D_{\sigma,P}(\xi^{(j)}) \right|
\end{aligned}
$$

More Observations

We have:

- $|D_{\sigma,P}(x)| = O(\sqrt{\log n} \cdot e^{-\Omega(\|x\|^2)}) = O(\sqrt{\log n})$ for all $x \in \mathbb{R}^d$
- For $x, y \in \mathbb{R}^d$,

$$\begin{aligned} & |D_{\sigma,P}(x) - D_{\sigma,P}(y)| \\ & \leq \left| \|y\|^2 - \|x\|^2 \right| |D_{\sigma,P}(x)| + 2 \sum_{j=1}^{d} |y_j - x_j| \left| D_{\sigma,P}(\xi^{(j)}) \right| \end{aligned}$$

It implies:

- The slope of $D_{\sigma,P}$ is $\tilde{O}(\sqrt{\log n})$
- The cell width is $\tilde{\Omega}(1/\sqrt{\log n})$

More Observations

Also, we have:

- $|D_{\sigma,P}(x)| = O(\sqrt{\log n} \cdot e^{-\Omega(\|x\|^2)})$ for all $x \in \mathbb{R}^d$

More Observations

Also, we have:

- $|D_{\sigma,P}(x)| = O(\sqrt{\log n} \cdot e^{-\Omega(\|x\|^2)})$ for all $x \in \mathbb{R}^d$

It implies:

- $|D_{\sigma,P}(x)| = O(e^{-\Omega(\|x\|^2)})$ when $\|x\|_\infty = \Omega(\sqrt{\log \log n})$
- The grid width is $O(\sqrt{\log \log n})$

More Observations

From the observations:

- The slope of $D_{\sigma,P}(x)$ is ~~$O(n)$~~ $\tilde{O}(\sqrt{\log n})$
- $|D_{\sigma,P}(x)| = O(e^{-\Omega(\|x\|^2)})$ when $\|x\|_\infty =$ ~~$\Omega(\sqrt{\log n})$~~ $\tilde{\Omega}(\sqrt{\log \log n})$

It implies:

- The cell width is ~~$\Omega(1/n)$~~ $\tilde{\Omega}(1/\sqrt{\log n})$
- The grid width is ~~$O(\sqrt{\log n})$~~ $\tilde{O}(\sqrt{\log \log n})$

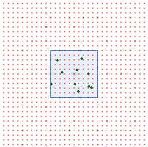
More Observations

From the observations:

- The slope of $D_{\sigma,P}(x)$ is ~~$O(n)$~~ $\tilde{O}(\sqrt{\log n})$
- $|D_{\sigma,P}(x)| = O(e^{-\Omega(\|x\|^2)})$ when $\|x\|_\infty =$ ~~$\Omega(\sqrt{\log n})$~~ $\tilde{\Omega}(\sqrt{\log\log n})$

It implies:

- The cell width is ~~$\Omega(1/n)$~~ $\tilde{\Omega}(1/\sqrt{\log n})$
- The grid width is ~~$O(\sqrt{\log n})$~~ $\tilde{O}(\sqrt{\log\log n})$



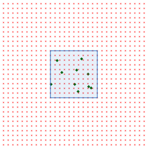
More Observations

From the observations:

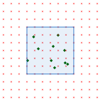
- The slope of $D_{\sigma,P}(x)$ is ~~$O(n)$~~ $\tilde{O}(\sqrt{\log n})$
- $|D_{\sigma,P}(x)| = O(e^{-\Omega(\|x\|^2)})$ when $\|x\|_\infty =$ ~~$\Omega(\sqrt{\log n})$~~ $\tilde{\Omega}(\sqrt{\log \log n})$

It implies:

- The cell width is ~~$\Omega(1/n)$~~ $\tilde{\Omega}(1/\sqrt{\log n})$
- The grid width is ~~$O(\sqrt{\log n})$~~ $\tilde{O}(\sqrt{\log \log n})$



$\Rightarrow$



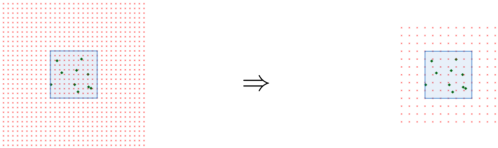
More Observations

From the observations:

- The slope of $D_{\sigma,P}(x)$ is ~~$O(n)$~~ $\tilde{O}(\sqrt{\log n})$
- $|D_{\sigma,P}(x)| = O(e^{-\Omega(\|x\|^2)})$ when $\|x\|_\infty =$ ~~$\Omega(\sqrt{\log n})$~~ $\tilde{\Omega}(\sqrt{\log \log n})$

It implies:

- The cell width is ~~$\Omega(1/n)$~~ $\tilde{\Omega}(1/\sqrt{\log n})$
- The grid width is ~~$O(\sqrt{\log n})$~~ $\tilde{O}(\sqrt{\log \log n})$



Size of the grid: $O((\log n)^{O(d)})$

Improvement

We have:

- $\mathbf{Pr}[|D_{\sigma,P}(x)| > \alpha] < O(e^{-\Omega(\alpha^2)})$ for one $x \in \mathbb{R}^d$
- $|D_{\sigma,P}(x)|$ is small on x in a finite grid of size $O((\log n)^{O(d)})$ implies $|D_{\sigma,P}(x)|$ is small on $x \in \mathbb{R}^d$

Improvement

We have:

- $\mathbf{Pr}[|D_{\sigma,P}(x)| > \alpha] < O(e^{-\Omega(\alpha^2)})$ for one $x \in \mathbb{R}^d$
- $|D_{\sigma,P}(x)|$ is small on x in a finite grid of size $O((\log n)^{O(d)})$ implies $|D_{\sigma,P}(x)|$ is small on $x \in \mathbb{R}^d$

If we set $\alpha = O(\sqrt{\log\log n})$ then we have

$$\left|\sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}\right| = O(\sqrt{\log\log n} \cdot e^{-\Omega(\|x\|^2)})$$

for all $x \in \mathbb{R}^d$

Induction

If we have $|D_{\sigma,P}(x)| = O(\beta \cdot e^{-\Omega(\|x\|^2)})$ for all $x \in \mathbb{R}^d$

Induction

If we have $|D_{\sigma,P}(x)| = O(\beta \cdot e^{-\Omega(\|x\|^2)})$ for all $x \in \mathbb{R}^d$

$$\Downarrow$$

- The slope of $D_{\sigma,P}(x)$ is $\tilde{O}(\beta)$
- $|D_{\sigma,P}(x)| = O(e^{-\Omega(\|x\|^2)})$ when $\|x\|_\infty = \Omega(\sqrt{\log \beta})$

Induction

If we have $|D_{\sigma,P}(x)| = O(\beta \cdot e^{-\Omega(\|x\|^2)})$ for all $x \in \mathbb{R}^d$

$$\Downarrow$$

- The slope of $D_{\sigma,P}(x)$ is $\tilde{O}(\beta)$
- $|D_{\sigma,P}(x)| = O(e^{-\Omega(\|x\|^2)})$ when $\|x\|_\infty = \Omega(\sqrt{\log \beta})$

$$\Downarrow$$

- The cell width is $\tilde{\Omega}(1/\beta)$
- The grid width is $\tilde{O}(\sqrt{\log \beta})$

Induction

If we have $|D_{\sigma,P}(x)| = O(\beta \cdot e^{-\Omega(\|x\|^2)})$ for all $x \in \mathbb{R}^d$

$$\Downarrow$$

- The slope of $D_{\sigma,P}(x)$ is $\tilde{O}(\beta)$
- $|D_{\sigma,P}(x)| = O(e^{-\Omega(\|x\|^2)})$ when $\|x\|_\infty = \Omega(\sqrt{\log \beta})$

$$\Downarrow$$

- The cell width is $\tilde{\Omega}(1/\beta)$
- The grid width is $\tilde{O}(\sqrt{\log \beta})$

$$\Downarrow$$

Size of the grid: $O(\beta^{O(d)})$

Induction

If we have $|D_{\sigma,P}(x)| = O(\beta \cdot e^{-\Omega(\|x\|^2)})$ for all $x \in \mathbb{R}^d$

$$\Downarrow$$

- The slope of $D_{\sigma,P}(x)$ is $\tilde{O}(\beta)$
- $|D_{\sigma,P}(x)| = O(e^{-\Omega(\|x\|^2)})$ when $\|x\|_\infty = \Omega(\sqrt{\log \beta})$

$$\Downarrow$$

- The cell width is $\tilde{\Omega}(1/\beta)$
- The grid width is $\tilde{O}(\sqrt{\log \beta})$

$$\Downarrow$$

Size of the grid: $O(\beta^{O(d)})$

$$\Downarrow$$

If we set $\alpha = O(\sqrt{\log \beta})$ in Banaszczyk's Theorem then we have $|D_{\sigma,P}(x)| = O(\sqrt{\log \beta} \cdot e^{-\Omega(\|x\|^2)})$ for all $x \in \mathbb{R}^d$

Induction

After performing $\log^* n$ inductive steps, we have

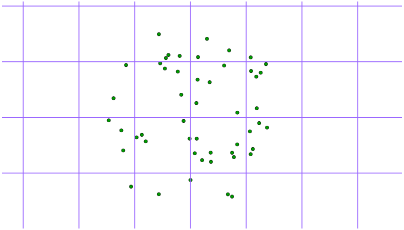
$$\left|\sum_{p\in P}\sigma(p)e^{-\|x-p\|^2}\right| = O(\sqrt{\log^* n \log\log^* n}\cdot e^{-\Omega(\|x\|^2)})$$

for all $x \in \mathbb{R}^d$ if P lies inside a ℓ_∞ ball of radius 1

Final Bound

Removing the assumption of P being inside a ℓ_∞ ball of radius 1:

- partition the input P
- run our algorithm on each ℓ_∞ ball of radius 1
- an extra constant factor depending on d in the final bound



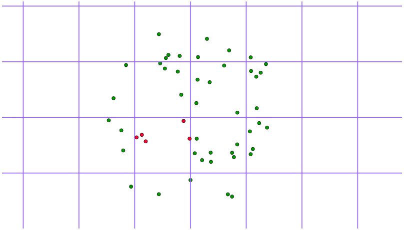
Our Result: $\left|\sum_{p\in P}\sigma(p)e^{-\|x-p\|^2}\right| = O(\sqrt{\log^* n \log\log^* n})$

Size of coreset: $O(\frac{1}{\varepsilon}\sqrt{\log^* \frac{1}{\varepsilon} \log\log^* \frac{1}{\varepsilon}})$ when d is constant

Final Bound

Removing the assumption of P being inside a ℓ_∞ ball of radius 1:

- partition the input P
- run our algorithm on each ℓ_∞ ball of radius 1
- an extra constant factor depending on d in the final bound



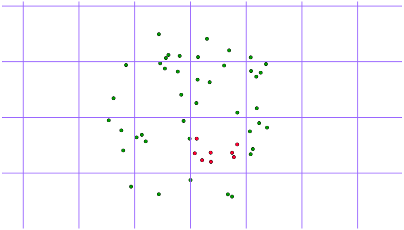
Our Result: $\left|\sum_{p \in P} \sigma(p) e^{-\|x-p\|^2}\right| = O(\sqrt{\log^* n \log\log^* n})$

Size of coreset: $O(\frac{1}{\varepsilon}\sqrt{\log^* \frac{1}{\varepsilon} \log\log^* \frac{1}{\varepsilon}})$ when d is constant

Final Bound

Removing the assumption of P being inside a ℓ_∞ ball of radius 1:

- partition the input P
- run our algorithm on each ℓ_∞ ball of radius 1
- an extra constant factor depending on d in the final bound



Our Result: $\left|\sum_{p\in P} \sigma(p) e^{-\|x-p\|^2}\right| = O(\sqrt{\log^* n \log\log^* n})$

Size of coreset: $O(\frac{1}{\varepsilon}\sqrt{\log^* \frac{1}{\varepsilon} \log\log^* \frac{1}{\varepsilon}})$ when d is constant

Conclusion

Find a ε-coreset $Q \subset P$ s.t.

$$\max_{x \in \mathbb{R}^d} \left| \overline{\mathcal{G}}_P(x) - \overline{\mathcal{G}}_Q(x) \right| \leq \varepsilon$$

Best known result for size of Q:

d	Upper	Lower	
1	$1/\varepsilon$	$1/\varepsilon$	
constant	$\frac{1}{\varepsilon}\sqrt{\log^* \frac{1}{\varepsilon} \log \log^* \frac{1}{\varepsilon}}$		**New result**
any	$\sqrt{d}/\varepsilon \cdot \sqrt{\log \frac{1}{\varepsilon}}$	$\sqrt{d}/\varepsilon$	
$\geq \frac{1}{\varepsilon^2}$	$1/\varepsilon^2$	$1/\varepsilon^2$	

Thank you