Geometry Reading Group – ε-nets and their applications

Michael P. Kim

12 July 2016

1 Outline

ε-nets refer to a number of similar mathematical structures that are useful for approximating sets of points, and more specifically, metric spaces.

Plan:

- Probabilistic notion of approximation
- Geometric notion of approximation
  - A framework for approximation algorithms
  - Net-Trees

These notes follow Ch. 12 of [Har06] closely, and then draw from [HM14, HR15].

2 Glossary of definitions

Definition 2.1 (Range Space). *A* range space (X, R) *consists of a set of points X and a family R of subsets of X.*

Definition 2.2 (Projection). *For a subset $A \subseteq X$, we say that the* projection *of a range space (X, R) onto A is $P_R(A) = \{r \cap A | r \in R\}$.*

Definition 2.3 (Shattering). *A set of points $A \subseteq X$ is* shattered *by a set of ranges R, if $P_A(R)$ contains all subsets of A. If A is finite, then this is equivalent to $|P_A(R)| = 2^{|A|}$.*

Definition 2.4 (VC-Dimension). *The* VC-Dimension *of a range space (X, R) is the maximum cardinality subset of X that is shattered by R. If arbitrarily large finite subsets of X can be shattered by R, then the VC-dim is ∞.*

Definition 2.5 (Doubling Dimension). *The* doubling constant c_2 *of a metric space M is the maximum, over all balls b of radius r, of the number of balls of radius $r/2$ needed to cover b. The* doubling dimension $d_2 = \log c_2$.

3 Probabilistic ε-nets

First, we present the weakest notion of an ε-net.

Definition 3.1 (Probabilistic ε-Net)**.** *Let (X, R) be a range space and $A \subseteq X$ be finite. We say that $N \subseteq A$ is an ε-net for A if for all ranges $r \in R$,*

$$|r \cap A| \geq \varepsilon |A| \implies |r \cap N| \geq 1.$$

That is, an ε-net N mimics A in terms of existence of points on highly-occupied ranges of A.

Useful for algorithmic applications as well as analysis:

- Sampling – select representative points where exact identity is not important
- Union bounding – transfer argument on continuous space to representative finite net

Theorem 3.1 (ε-net Theorem)**.** *Let (X, R) be a range space of VC-dim d and $A \subseteq X$ be finite. For $0 < \varepsilon, \delta < 1$, construct $N \subseteq A$ by sampling $m_{\varepsilon,\delta}$ points independently uniformly at random from A. N is an ε-net for A with probability $\geq 1 - \delta$ provided*

$$m_{\varepsilon,\delta} \geq O\left(\frac{1}{\varepsilon}\log\frac{1}{\delta} + \frac{d}{\varepsilon}\log\frac{d}{\varepsilon}\right).$$

Proof Idea:
Let N, M be sets of m points randomly subsampled from A as in the theorem statement.

- Define the event $E_\varnothing = \{\exists r \in R : |r \cap A| \geq \varepsilon |A| \wedge r \cap N = \varnothing\}$
- Define the event $E_{\varnothing,1/2} = \left\{\exists r \in R : |r \cap A| \geq \varepsilon |A| \wedge r \cap N = \varnothing \wedge |r \cap M| \geq \frac{\varepsilon m}{2}\right\}$

Claim: $\mathbf{Pr}[E_{\varnothing,1/2}] \leq \mathbf{Pr}[E_\varnothing] \leq 2\,\mathbf{Pr}[E_{\varnothing,1/2}]$.
(First, obvious; Second, by Chebyshev's Inequality to bound $\mathbf{Pr}[E_{\varnothing,1/2} \big| E_\varnothing] \leq 1/2$.)

Key simplification: Ignore A.

- Define the event $E_0 = \left\{\exists r \in R : r \cap N = \varnothing \wedge |r \cap M| \geq \frac{\varepsilon m}{2}\right\}$

Sample $2m$ points Z, look at the probability that a random equipartition results in E_0. Let $k = |r \cap (N \cup M)|$.

$$\mathbf{Pr}[E_0] \leq |P_Z(R)| \cdot \mathbf{Pr}[r \cap N = \varnothing \big| k \geq \frac{\varepsilon m}{2}] \leq |P_Z(R)| \frac{\binom{2m-k}{m}}{\binom{2m}{m}} \leq |P_Z(R)| \cdot 2^{\varepsilon m/2}$$

Using Sauer's Lemma to bound $|P_Z(R)| \leq |R| \leq \sum_{i=0}^{d} \binom{2m}{i}$ and plugging in the asserted values for $m_{\varepsilon,\delta}$, we obtain the theorem.

4 Geometric ε-nets

For geometric applications, a more geometric concept is useful.

Definition 4.1 (Geometric ε-net)**.** *Let (X, d) be a metric space and $\varepsilon > 0$. We say that $N \subseteq X$ is an* ε-net *of X if*

(i) Packing – for all $x \neq y \in N$, $d(x, y) \geq \varepsilon$, and

(ii) Covering – for all $x \in X$, $\min_{y \in N} d(x, y) < \varepsilon$.

Simple Construction for ε-net for finite metrics:

- Initialize all $x \in X$ to be unmarked
- $N \leftarrow \varnothing$
- While there is some $x_u \in X$ that remains unmarked:
  - $N \leftarrow N \cup \{x_u\}$
  - mark x_u and all x' where $d(x_u, x') < \varepsilon$

Fact: Using hasing and geometric tricks, this construction can be performed in linear time in the size of the point set.

4.1 Approximate Optimization By "Net-and-Prune" Framework

[HR15] gives a general framework for constructing linear-time approximation algorithms for geometric problems. We will consider the k-center problem as a case study.

k-Center Problem: Let X be a set of points in $\mathbb{R}^D$ and $k > 0$ be some integer. Compute a subset $K \subseteq X$ where $|K| = k$ according to the following objective:

$$f(X, k) = \max_{x \in X} \min_{c \in K} d(x, c)$$

where $d : \mathbb{R}^D \to \mathbb{R}$ is the Euclidean distance over $\mathbb{R}^D$. We denote the optimal solution as

$$f^*(X, k) \triangleq \min_{\substack{K \subseteq X \\ |K| = k}} \max_{x \in X} \min_{c \in K} d(x, c).$$

The k-center problem exhibits three important properties that makes it a candidate for the net-and-prune framework.

- c-Approximate Decider: There exists a linear-time decider that, given (X, k) and some $\varepsilon > 0$ decides whether: (i) $f^*(X, k) \in [\alpha, c\alpha]$ for some $\alpha \in \mathbb{R}$, (ii) $f^*(X, k) < \varepsilon$, or (iii) $f^*(X, r) > \varepsilon$.
- Lipschitz: Suppose X_Δ is a Δ-drift of X (that is, there is mapping $q : X \to X_\Delta$, such that for all $x \in X$, $d(x, q(x)) \leq \Delta$). Then $|f^*(X, k) - f^*(X_\Delta, k)| \leq 2\Delta$.

- Invariant under Pruning: Let $X^{<\varepsilon} \subseteq X$ be the set of points, whose nearest neighbor in X is closer than ε. If $f^*(X,k) < \varepsilon$, then $f^*(X,k) = f^*(X^{<\varepsilon}, k^{<\varepsilon})$ where $k^{<\varepsilon} = k - |X \setminus X^{<\varepsilon}|$.

These properties are not difficult to verify; we call the framework "net-and-prune" because the c-approximate decider can frequently be implemented using an ε-net.

Claim: Let N be an ε-net for X.

- If $|N| \leq k$, then $f^*(X,k) < \varepsilon$. (implied by definition of ε-net)
- If $|N| > k$, then $f^*(X,k) \geq \varepsilon/2$. (consider $k+1$ points, all separated by ε)

Description of $(4+\delta)$-Approximate Decider:
Given (X,k) and ε, construct an ε-net N for X.

- If $|N| \leq k$, then return $f^*(X,k) \leq \varepsilon$.

Construct a $(2+\delta/2)\varepsilon$-net N' for X.

- If $|N'| \leq k$, then return $f^*(X,k) \in [\varepsilon/2, (2+\delta/2)\varepsilon]$.
- Else, return $f^*(X,k) > \varepsilon$.

Lipschitz: The distance between two points can increase by at most 2Δ in a Δ-drift of the original point set.

Invariant under Pruning: If $f^*(X,k) < \varepsilon$, then any points whose nearest-neighbor is $\geq \varepsilon$ must be a center of its own.

Algorithm:
k-center(X,k):

- Randomly sample $x \in X$
- $\varepsilon \leftarrow \min_{y \neq x \in X} d(x,y)$
- Run decider for $\langle (X,k), \varepsilon \rangle$ and $\langle (X,k), c_0\varepsilon \rangle$ (for some sufficiently large constant $c_0 > 37$)
- if either run of the decider finds a range such that $f^*(X,k) \in [x,y]$, return $f^*(X,k) \in [x/2, 2y]$
- else if $\varepsilon < f^*(X,k) < c_0\varepsilon$, then return $f^*(X,k) \in [\varepsilon/2, 2c_0\varepsilon]$
- else if $f^*(X,k) < \varepsilon$, *(prune)* return k-center$(X^{<\varepsilon}, k^{<\varepsilon})$
- else if $f^*(X,k) > c_0\varepsilon$, *(net)* compute a 3ε-net N of X and return k-center(N,k)

Proof Idea:
Runtime: In every recursive call, we throw away a constant fraction of the input in expectation.
Correctness: Either we prune (and the objective is not changed), or we net. The main crux of the proof is showing that the net radius is always significantly smaller than the objective value, and thus, the quality of approximation does not degrade rapidly.

Using a gridding technique and a standard greedy algorithm, one can turn any c-approximate solution to the k-center problem to a 2-approximate solution (see [Har04], Lemma 6.5).

4.2 Hierarchical Nets for Doubling Metrics

For a given metric space (X, d), we use $b_r(x)$ to denote the ball of radius r according to d surrounding some $x \in X$; that is $b_r(x) = \{x' \in X \,|\, d(x, x') \leq r\}$.

In this section, we will discuss the following extension of ε-nets.

Definition 4.2 (Net-tree)**.** *Let (X, d) be a metric space and let $P \subseteq X$ be finite. A net-tree of P is a tree T whose leaves represent P. For each internal node $v \in T$, we associate the following:*

- $P_v \subseteq P$ *– the set of leaves in P of the subtree rooted at v*
- p_v *– the parent of v in T (not defined for the root)*
- rep_v *– some $v \in P_v$*
- ℓ_v *– the level of v in the tree, satisfying $\ell_v < \ell_{p_v}$; ℓ_x for $x \in P$ defined to be $-\infty$.*

For sufficiently large $\tau = 11$, we require the following properties to be maintained:

(i) Packing – for every non-root $v \in T$:

$$b(\mathrm{rep}_v, \Delta_p) \cap P \subseteq P_v$$

for $\Delta_p = \frac{\tau-5}{2(\tau-1)} \cdot \tau^{\ell_{p_v}-1}$

(ii) Covering – for every $v \in T$:

$$P_v \subseteq b(\mathrm{rep}_v, \Delta_c)$$

for $\Delta_c = \frac{2\tau}{\tau-1} \cdot \tau^{\ell_v}$

(iii) Inheritance – for every non-leaf $v \in T$, there exists some $u \in T$ such that $v = p_u$ and $\mathrm{rep}_v = \mathrm{rep}_u$.

Constructing net-trees: Consider the following greedy process: choose an arbitrary point x_1 to start; add it to the list of centers C. Then, choose x_{i+1} to be

$$x_{i+1} \leftarrow \operatorname*{argmax}_{x \in X \setminus C} \min_{c \in C} d(x, c).$$

We define $r_k = \min_{1 \leq i,j \leq k+1} d(x_i, x_j)$.

In this order, we process the points $x_k \in P$. Given the tree consisting of the first $k-1$ points, $T^{(k-1)}$, we add x_k as a leaf.

- Let $\ell = \lceil \log_\tau r_{k-1} \rceil$
- Let $x^{(*)}$ be the nearest-neighbor of x_k such that $\ell_{x^{(*)}} > \ell$. Let $u = p_{x^{(*)}}$.
- If $\ell_u > \ell$, create a new internal node v such that $\mathrm{rep}_v = x^{(*)}$. Make $x^{(*)}$ and x_k children of v and v a child of u.
- Else, add x_k as a child of u.

Claim: The above process returns a net-tree.

4.2.1 Approximate Nearest Neighbor Search

We will design two different weak ANN structures that can be used in combination to yield a $(1+\varepsilon)$-ANN structure.

Low-spread: Suppose we're given a net-tree T of P, a query $q \in X$, and some $u \in T$ such that $d(\mathrm{rep}_u, q) \leq 5 \cdot \tau^\ell$, OR $x^{(*)} \in P_u$, where $x^{(*)}$ is the nearest-neighbor of q in P.

Algorithm:

- Construct $A_\ell \leftarrow \left\{ v \in T \middle| \ell(v) \leq \ell(u) \leq \ell(p_v) \wedge d(\mathrm{rep}_u, \mathrm{rep}_v) \leq 13\tau^{\ell(u)} \right\}$
- Construct A_{i-1} from A_i as follow:
  - Compute $d_0 \leftarrow \min_{w \in A_i} d(\mathrm{rep}_w, q)$
  - Replace all vertices in A_i with their children, but only keep a vertex $v \in T$ if $d(\mathrm{rep}_v, q) \leq d_0 + \frac{2\tau}{\tau-1}\tau^{i-1}$

Low-quality BST: Suppose we just want to find a $2n$-ANN. We proceed by building a BST, where the left search tree is constructed for $P \cap b_{r_v}(p_v)$ and left search tree is constructed for $P \setminus b_{r_v}(p_v)$ for some specially selected v.

Combination:

- Use the $2n$-ANN structure to find a close point x_0
- Find an ancestor of x_0 in net-tree T that is at an appropriate level
- Run the search on the net-tree, knowing that the spread is sufficiently bounded

References

[Har04] Sariel Har-Peled. Clustering motion. *Discrete & Computational Geometry*, 2004.

[Har06] Sariel Har-Peled. Geometric approximation algorithms. 2006.

[HM14] Sariel Har-Peled and Manor Mendel. Fast construction of nets in low-dimensional metrics and their applications. *SICOMP*, 2014.

[HR15] Sariel Har-Peled and Benjamin Raichel. Net and prune: A linear time algorithm for euclidean distance problems. *JACM*, 2015.