COMS E6998-9 F15

Lecture 15:
Least Square Regression
Metric Embeddings

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Administrivia, Plan

- PS2:
  - Pick up after class
- 120->144 auto extension

- Plan:
  - Least Squares Regression (finish)
  - Metric Embeddings
    - “reductions for distances”

Least Square Regression

- Problem:
  - $argmin_x||Ax - b||$
  - where A is $n \times d$ matrix
  - b is a vector of dimension n
  - $n \gg d$
- Usual (exact) solution:
  - Perform SVD (singular value decomposition)
  - Takes $O(nd^{\omega-1}) \approx O(nd^{1.373})$ time
- Faster?

Approximate LSR

- Approximate solution x':
  - $||Ax' - b|| \leq (1+\epsilon)||Ax^* - b||$
    - Where x^* optimal solution
- Tool: dimension reduction for subspaces!
  - A map $\Pi: \Re^n \to \Re^k$ is (d, ϵ, δ)-subspace embedding if
    - For any linear subspace $P \subset \Re^n$ of dimension d, we have that

$$\Pr_{\Pi}\left[\forall p \in P : \frac{||\Pi(p)||}{||p||} \in (1-\epsilon, 1+\epsilon)\right] \geq 1-\delta$$

  - PS3-2: usual dimension reduction implies $(d, \epsilon, 0.1)$ for target dimension $k = O(d/\epsilon^2)$

Approximate Algorithm

- Algorithm:
  - Let Π be $(d+1, \epsilon, 0.1)$-subspace embedding
  - Solve $x' = \text{argmin}\ ||\Pi A x - \Pi b||$
- Theorem: $||Ax' - b|| \leq (1+3\epsilon)||Ax^* - b||$
- Proof:
  - $||\Pi A x - \Pi b|| = ||\Pi(Ax - b)||$
  - $P = \{Ax - b \mid x \in \Re^d\}$ is a (subset of a) $d+1$ dimensional linear subspace of $\Re^n$
  - Hence Π preserves the norm of all $Ax - b$
    - Up to $1+\epsilon$ approximation each
    - with 90% probability (overall)

1) $||\Pi A x^* - \Pi b|| \leq (1+\epsilon)||Ax^* - b||$
2) $||\Pi A x - \Pi b|| \geq (1-\epsilon)||Ax - b||$ for any x

Hence $||Ax' - b|| \leq \frac{1+\epsilon}{1-\epsilon}||Ax^* - b||$

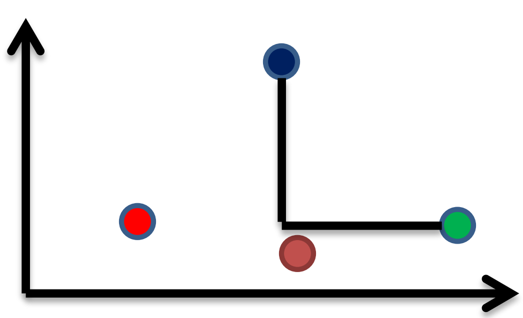
Approx LSR algorithms

- Time:
  - $O(nd^{\omega-1}) \Rightarrow O(kd^{\omega-1}) = O_\epsilon(d^\omega)$
  - Plus time to multiply by Π: $O(ndk) = O_\epsilon(nd^2)$
    - This is worse than before in fact…
- Can apply Fast dimension reduction!
  - Reduce time to:
    - $O\big(d \cdot (n \cdot \log n + d^3)\big) = O(nd \cdot \log n + d^4)$
    - First term near optimal
- Can do even faster:
  - Exist Π with 1 non-zero/column with $k = O(d^2/\epsilon^2)$
    - Exactly the one from problem 1 on PS2 !
  - Time becomes: $O_\epsilon(nnz(A) + d^3)$
- [Sarlos'06, Clarkson-Woodruff'13, Meng-Mahoney'13, Nelson-Nguyen'13]

Metric embeddings

Definition by example

- Problem: Compute the diameter of a set S, of size n, living in d-dimensional ℓ_1^d
  - Say, for $d = 2$
- Trivial solution: $O(dn^2)$ time
- Will see: $O(2^d n)$ time
- Algorithm has two steps:
  1. Map $f\colon \ell_1^d \rightarrow \ell_\infty^k$, where $k = 2^d$ such that, for any $x, y \in \ell_1^d$
     - $||x - y||_1 = ||f(x) - f(y)||_\infty$
  2. Solve the diameter problem in ℓ_∞ on set $f(S)$



Step 1: Map from ℓ_1 to ℓ_∞

- Want map $f: \ell_1 \to \ell_\infty$ such that for $x, y \in \ell_1$
  - $\|x-y\|_1 = \|f(x)-f(y)\|_\infty$
- Define $f(x)$ as follows:
  - 2^d coordinates indexed by $b = (b_1 b_2 \ldots b_d)$ (binary representation)
  - $f(x)_b = \sum_i (-1)^{b_i} \cdot x_i$
- Claim: $\|f(x)-f(y)\|_\infty = \|x-y\|_1$

$$\begin{aligned}\|f(x)-f(y)\| &= \max_b \sum_i (-1)^{b_i} \cdot (x_i - y_i) \\ &= \textstyle\sum_i \max_{b_i} (-1)^{b_i} (x_i - y_i) \\ &= \textstyle\sum_i |x_i - y_i| \\ &= ||x-y||_1\end{aligned}$$

Step 2: Diameter in ℓ_∞

- Claim: can compute diameter of n points living in ℓ_∞^k in $O(nk)$ time.

- Proof:

$$\text{diameter}(S) = \max_{p,q \in S} ||p - q||_\infty$$
$$= \max_{x,y \in S} \max_b |p_b - q_b|$$
$$= \max_b \max_{p,q \in S} |p_b - q_b|$$
$$= \max_b (\max_{p \in S} p_b - \min_{q \in S} q_b)$$

- Hence, can compute in $O(k \cdot n)$ time.
- Combining the two steps, we have $O(2^d \cdot n)$ time for computing diameter in ℓ_1^d

General Theory: embeddings

- General motivation: given distance (metric) M, solve a computational problem P under M

Hamming distance, ℓ_1

Euclidean distance (ℓ_2)

Edit distance between two strings

Earth-Mover (transportation) Distance

Compute distance between two points

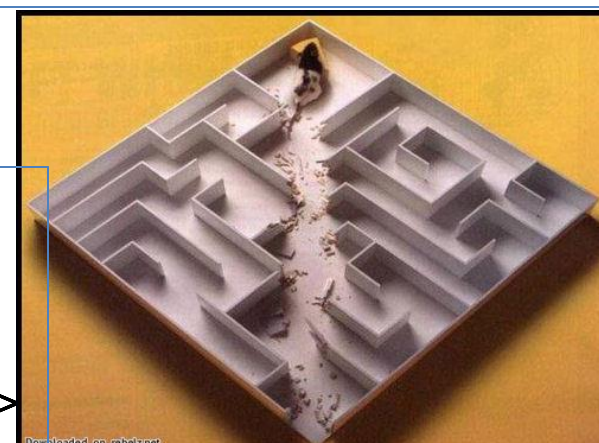
Nearest Neighbor Search

Diameter/Close-pair of set S

Clustering, MST, etc

f

Reduce problem
<P under hard metric>
to
<P under simpler metric>



Embeddings: landscape

- **Definition**: an embedding is a map $f : M \to H$ of a metric (M, d_M) into a host metric (H, ρ_H) such that for any $x, y \in M$:

  $$d_{M(x,y)} \leq \rho_H(f(x), f(y)) \leq D \cdot d_M(x, y)$$

  where D is the distortion (approximation) of the embedding f.

- Embeddings come in all shapes and colors:
  - Source/host spaces M, H
  - Distortion D
  - Can be randomized: $\rho_H(f(x), f(y)) \approx d_M(x, y)$ with $1 - \delta$ probability
  - Time to compute $f(x)$
- Types of embeddings:
  - From norm to the same norm but of *lower dimension* (dimension reduction)
  - From one norm (ℓ_2) into another norm (ℓ_1)
  - From non-norms (edit distance, Earth-Mover Distance) into a norm (ℓ_1)
  - From given finite metric (shortest path on a planar graph) into a norm (ℓ_1)
  - H not a metric but a computational procedure ← sketches

ℓ_2 into ℓ_1

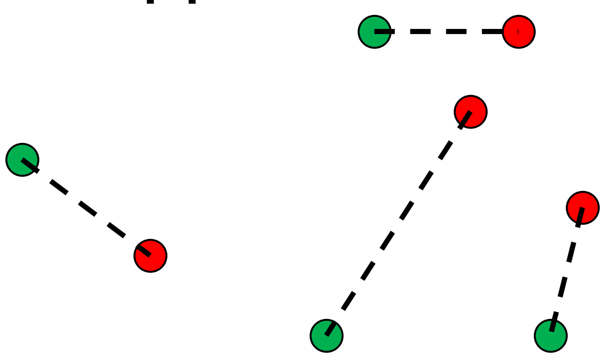
- Theorem: can embed ℓ_2^d into ℓ_1^k for $k = O\left(\frac{d}{\epsilon^2}\right)$ and distortion $1+\epsilon$
  - Map: $F(x) = \frac{1}{k}Gx$ for G = Gaussian $k \times d$
- Proof:
  - Idea similar to dimension reduction in ℓ_2:
  - Claim: for any points $x, y \in \Re^d$, let $\delta = ||x-y||_2$, then:
    - $\Pr\left[\frac{||F(x)-F(y)||_1}{\delta} \in (1-\epsilon, 1+\epsilon)\right] \geq 1 - e^{-\Omega(\epsilon^{-2}k)}$
  - Proof:
    - $F(x) - F(y) = \frac{1}{k}G(x-y)$
    - Distributed as $\frac{1}{k}(g_1\delta, g_2\delta, \dots, g_k\delta)$
    - Hence $||F(x) - F(y)||_1 = \delta \cdot \frac{1}{k}\sum_i |g_i|$
      - (in dimension reduction we had $\frac{1}{k}\sum_i g_i^2$)
    - Also can prove that: $\frac{1}{k}\sum_i |g_i| = 1 \pm \epsilon$ with probability at least $1 - e^{-\Omega(\epsilon^{-2}k)}$
  - Now apply the same argument as in subspace embedding to argue for the entire space $\Re^d$ as long as $k \geq \Omega(d/\epsilon^2)$
- Morale: ℓ_1 is at least as “large/hard” as ℓ_2

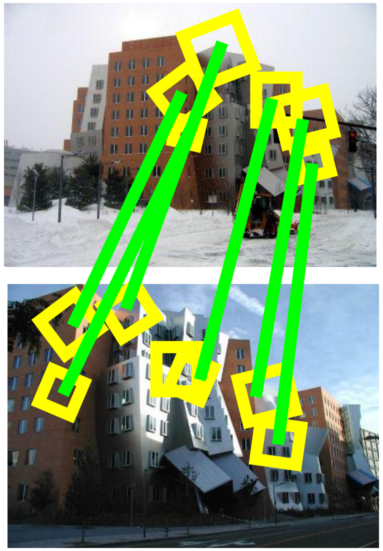
Converse?

- Can we embed ℓ_1 into ℓ_2 with good approximation?
  - No!
- [Enflo'69]: embedding $\{0,1\}^d$ into ℓ_2 (any dimension) must incur $\sqrt{d}$ distortion
- Proof:
  - Suppose f is the embedding of $\{0,1\}^d$ into ℓ_2
  - Two distributions over pairs of points $x, y \in \{0,1\}^d$:
    - F: x random and $y = x \oplus \bar{1}$
    - C: $x = y \oplus e_i$ for random y and index i
  - Two steps:
    - $E_C\left[\|x-y\|_1^2\right] \leq 1/d^2 \cdot E_F\left[\|x-y\|_1^2\right]$
    - $E_C\left[\|f(x)-f(y)\|_2^2\right] \geq 1/d \cdot E_F\left[\|f(x)-f(y)\|_2^2\right]$
      - (short diagonals)
  - Implies $\Omega(\sqrt{d})$ lower bound!
- Morale: ℓ_1 is strictly larger than ℓ_2 !

Other distances?

- E.g, Earth-Mover Distance
- Definition:
  - Given two sets A, B of points in a metric space
  - $EMD(A, B)$ = min cost bipartite matching between A and B
- Which metric space?
  - Can be plane, ℓ_2, ℓ_1…
- Applications in image vision





Images courtesy of Kristen Grauman