Topics in Algorithms 2007

Ramesh Hariharan

Random Projections

Solving High Dimensional Problems

- How do we make the problem smaller?
- Sampling, Divide and Conquer, what else?

Projections

- Can n points in m dimensions be projected to d<<m dimensions while maintaining geometry (pairwise distances)?

- Johnson-Lindenstrauss: YES

for $d \sim \log n/\varepsilon^2$, each distance stretches/contracts by only an $O(\varepsilon)$ factor

So an algorithm with running time f(n,d) now takes f(n,log n) and results don't change very much (hopefully!)

Which d dimensions?

- Any d coordinates?
- Random d coordinates?
- Random d dimensional subspace

Random Subspaces

- How is this defined/chosen computationally?

- How do we choose a random (spherically symmetric) line (1-d subspace)

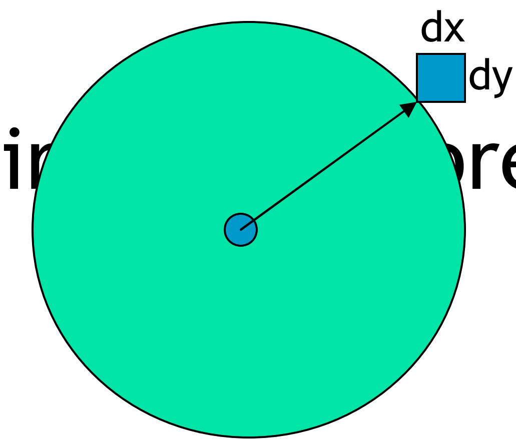
Random Subspaces

- We need to choose m coordinates
- Choices:
  - Independent Uniform distributions on cartesian coordinates
    - Not defined: infinite ranges
    - Not spherically symmetric: different axis segments of the same length map surface patches with different surface areas
  - Independent Uniform distributions on polar coordinates
    - Works in 2d
    - Does not work in higher dimensions!! Why? for a 3d sphere, half and not one-third the area is located within 30 degrees

Random Subspaces

- What works?
- Normal distribution on cartesian coordinates
- Choose independent random variables $X_1 \ldots X_m$ each $N(0,1)$

Why do Normals work?



- Take 2d: Which points on the ci re likely?

$$e^{-x^2/2}\, dx \quad X \quad e^{-y^2/2}\, dy = e^{-x^2/2-y^2/2}\, dydx$$

A random d-dim subspace

- How do we extend to d dimensions?
- Choose d random vectors

Choose independent random variables $X^i_1 \ldots X^i_m$ each $N(0,1)$, $i=1..d$

Distance preservation

- There are ${}^{n}C_2$ distances
- What happens to each after projection?
- What happens to one after projection; consider single unit vector along x axis

- Length of projection $\mathrm{sqrt}\left[\left(x^1{}_1 / l_1 \right)^2 + \ldots + \left(x^d{}_1 / l_d \right)^2 \right]$

Orthogonality

- Not exactly
- The random vectors aren't orthogonal
- How far away from orthogonal are they?

- In high dimensions, they are nearly orthogonal!!

Assume Orthogonality

- Assume orthogonality for the moment
- How do we determine bounds on the distribution of the projection length

$$\mathrm{sqrt}\left[\left(x^1{}_1 / l_1\right)^2 + \ldots + \left(x^d{}_1 / l_d\right)^2\right]$$

- Expected value of $\left[\left(x^1{}_1\right)^2 + \ldots + \left(x^d{}_1\right)^2\right]$ is d (by linearity of expectation)
- Expected value of each l_i^2 is m (by linearity of expectation)
- Roughly speaking, overall expectation is sqrt(d/m)
- A distance scales by sqrt(d/m) after projection in the "expected" sense; how much does it depart from this value?

What does Expectation give us

- But E(A/B) != E(A)/E(B)
- And even if it were, the distribution need not be tight around the expectation
- How do we determine tightness of a distribution?
- Tail Bounds for sums of independent random variables; summing gives concentration (unity gives strength!!)

Tail Bounds

- $P(|\Sigma^k X_i^2 - k| > \varepsilon k) < 2e^{-\varepsilon^2 k/4}$

$$\text{sqrt}\left[\left(x^1{}_1 / l_1\right)^2 + \ldots + \left(x^d{}_1 / l_d\right)^2\right]$$

- Each l_i^2 is within $(1 +/- \varepsilon)m$ with probability inverse exponential in $\varepsilon^2 m$
- $[(x^1{}_1)^2 + \ldots + (x^d{}_1)^2]$ is within $(1 +/- \varepsilon)d$ with probability inverse exponential in $\varepsilon^2 d$
- $\text{sqrt}[(x^1{}_1/l_1)^2 + \ldots + (x^d{}_1/l_d)^2]$ is within $(1 +/- O(\varepsilon))\ \text{sqrt}(d/m)$ with probability inverse exponential in $\varepsilon^2 d$ (by the union bound)

One distance to many distances

- So one distance D has length D (1 +/- O(ε)) sqrt(d/m) after projection with probability inverse exponential in $\varepsilon^2 d$
- How about many distances (could some of them go astray?)
- There are ${}^nC_2$ distances
- Each has inverse exponential in $\varepsilon^2 d$ probability of failure, i.e., stretching/compressing beyond (1 +/- O(ε)) sqrt(d/m)
- What is the probability of no failure? Choose d so that $e^{-\varepsilon^2 d/4}$

* ${}^nC_2$ is << 1 (union bound again), so d is $\Theta(\log n/ \varepsilon^2)$

Orthogonality

- What is the distribution of angles between two spherically symmetric vectors
- Tight around 90 degrees!!

Orthogonality

- Distribution for one vector X is $e^{-(X^TX)/2}\,dX$

- Invariant to rotation
  - Rotation is multiplication by square matrix R where $R^TR=I$
  - Take Y=RX so $R^TY=X$
  - Distribution of Y is $e^{-((RY)^T(RY)/2}\,dY = e^{-(Y^TY)/2}\,dY$

- Y is sph sym in the n-1 dim space ortho to X
  - Z= RX so that Z=10000..
  - RX, RY are sph sym
  - So RY comprises n independent N(0,1)'s
  - The portion of RY orthogonal to RX has n-1 independent N(0,1)'s

Orthogonality

- Start with v=10000…
- Projection on first vector has length X/L where X is N(0,1) and L is sqrt of sum of squares of N(0,1)
- Project v and all other vectors in the space ortho to the first vector
- Again rotate within this orthogonal space so v=sqrt[1- $(X/L_1)^2$] 0 0 0 0 0
- All other vectors are still sph sym in this ortho space
- So projection of v on second vector has length X'/L' * sqrt[1- $(X/L_1)^2$]
- And so on.. overall dampening is $(\mathrm{sqrt}[1-(X/L_1)^2])^d = (1-\log n/n)^{d/2}$ whp

Tail Bound Proof

Show $P(|\Sigma^k X_i^2 - k| > \varepsilon k) < 2e^{-\varepsilon^2 k/2}$

Two parts

- $P(\Sigma^k X_i^2 < (1-\varepsilon)k) < e^{-\varepsilon^2 k/2}$
- $P(\Sigma^k X_i^2 > (1+\varepsilon)k) < e^{-\varepsilon^2 k/2}$

Tail Bound Proof

$P(\Sigma^k X_i^2 < (1-\varepsilon)k) < e^{-\varepsilon^2 k/4}$

$= P(e^{-t\Sigma X^2} > e^{-t(1-\varepsilon)k})$:t>0

$\leq E(e^{-t\Sigma X^2})/e^{-t(1-\varepsilon)k}$:Markov's

$\leq E(e^{-tX^2})^k e^{t(1-\varepsilon)k}$:Independence

$\leq (2t+1)^{-k/2} e^{t(1-\varepsilon)k}$:Why?

Minimize for t and complete the proof??

Exercises

- Complete proof of tail bound
- Take a random walk in which you move left with prob p and right with prob 1-p; what can you say about your distance from the start after n steps?