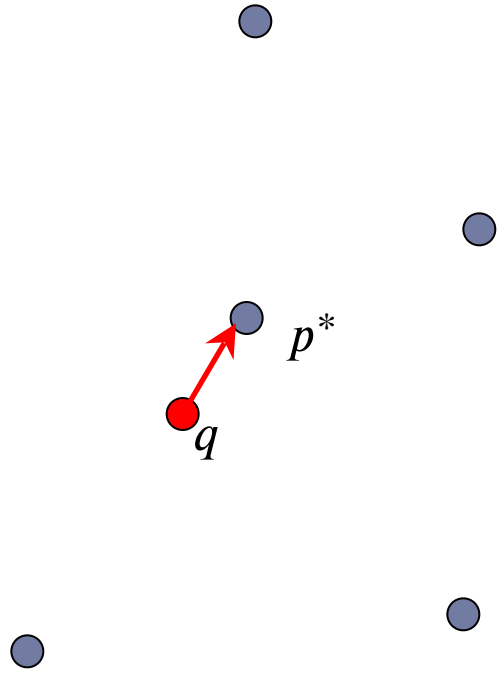
Learning to Hash Robustly, Guaranteed

Daniel Beaglehole (UCSD)

Joint with Alexandr Andoni (Columbia)

Nearest Neighbor Search (NNS)

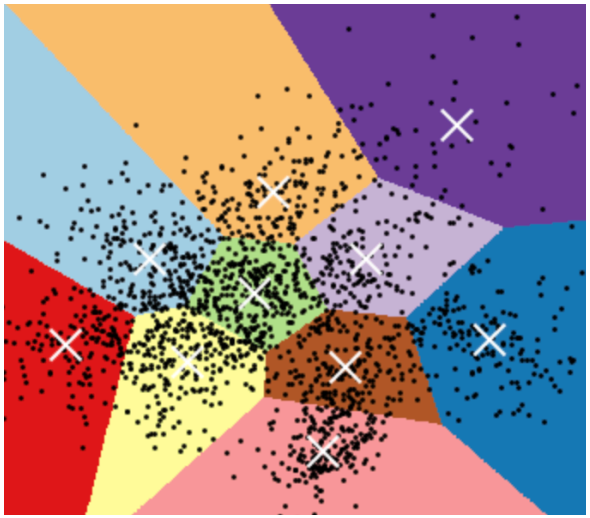
- Pre-process: a set P of points
- Query: given a query point q, report a point $p^* \in P$ with the smallest distance to q



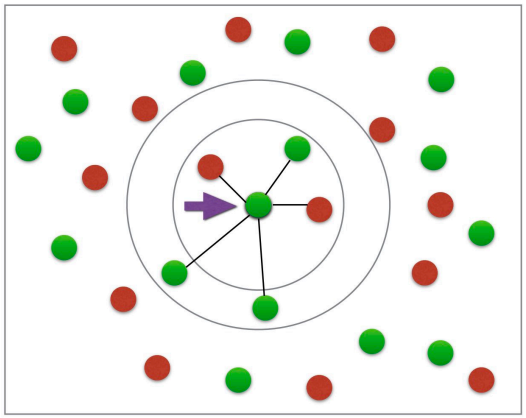
Motivation

Many applications for NNS across domains. For example,

Clustering



k-NN predictors



Approximate Nearest Neighbors Search

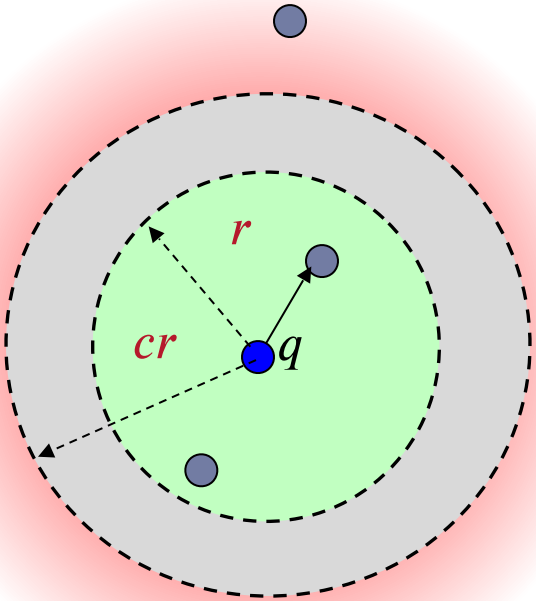
NNS is hard when $d > \log n$ (either exponential in dimension or linear in data size)

NNS

relax

c-approximate

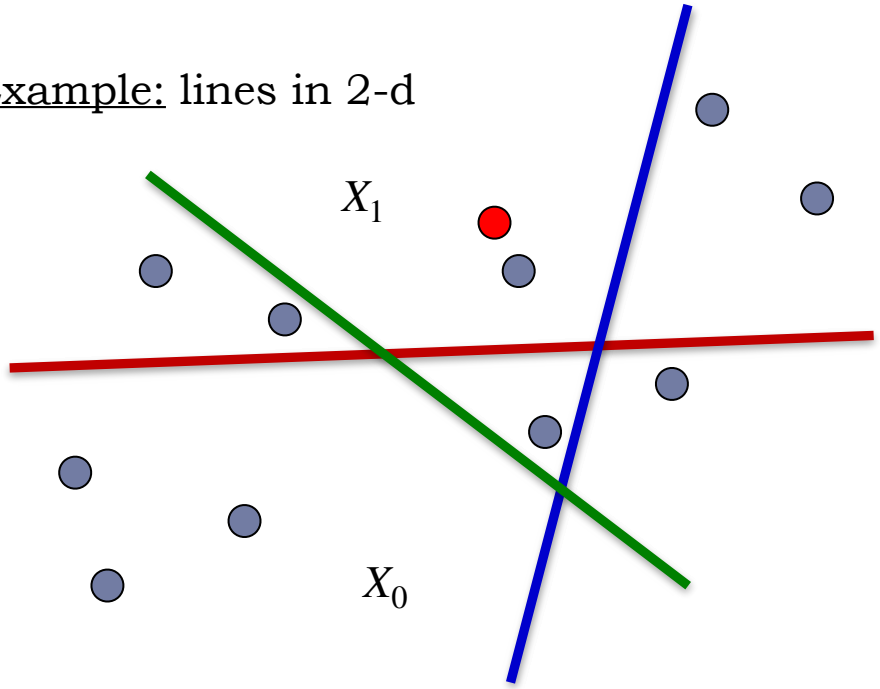
- r-near neighbor: given a query point q, report a point $p' \in P$ s.t. $\|p' - q\| \leq r$
  - as long as there is some point within distance r

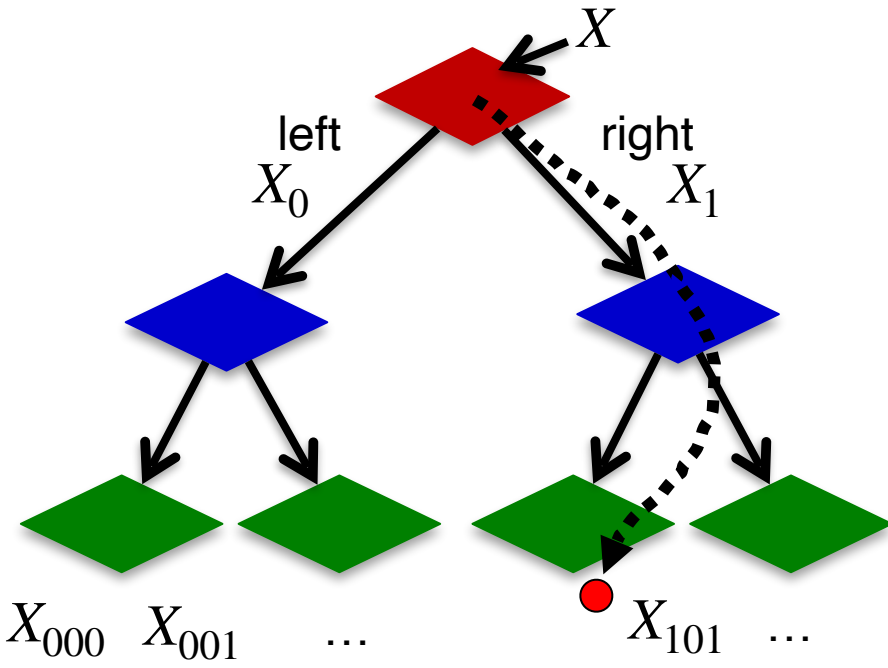


Locality-Sensitive Hashing (LSH)

- Spatial partitions (e.g. random hyperplanes)
- Worst-case theoretical guarantees
- Oblivious to the dataset

Example: lines in 2-d

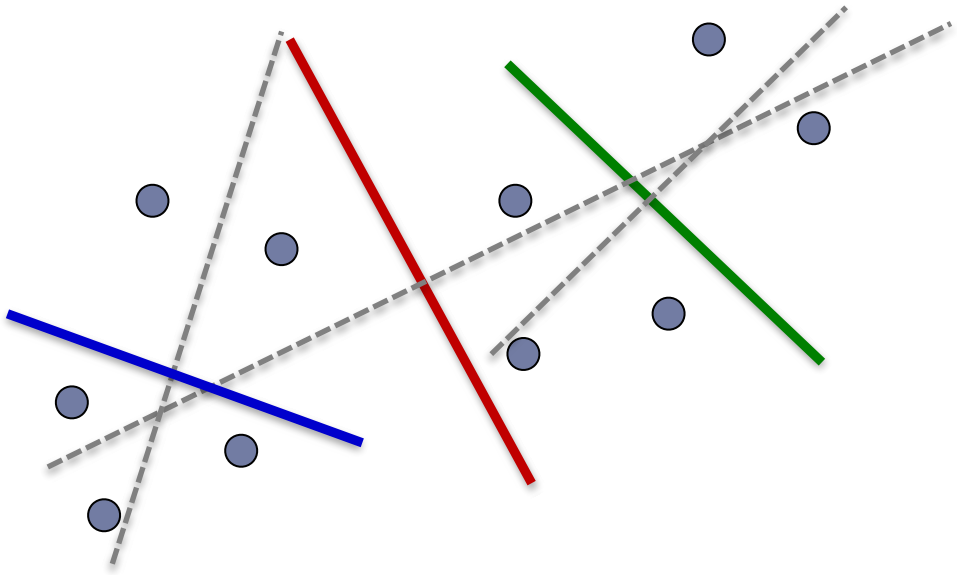




Only need $O(n^{1/c})$ time, to find an approximate near neighbor with high probability

vs Practice

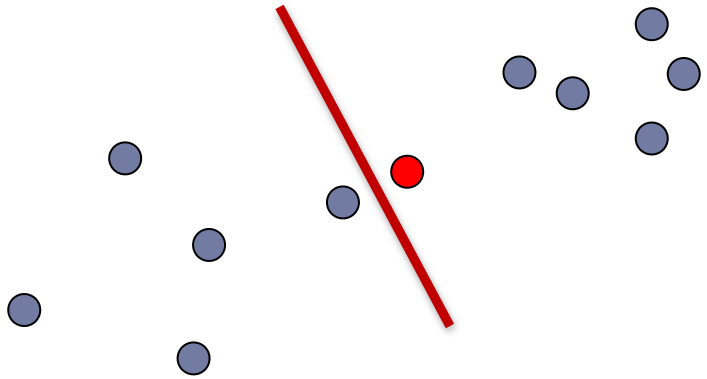
- *Data-dependent*: optimize the partition to *your* dataset
  - PCA-tree [S'91, M'01, VKD'09,...]



Practice vs Theory

- Data-dependent partitions often outperform random partitions on average
- But no guarantees (correctness or performance)

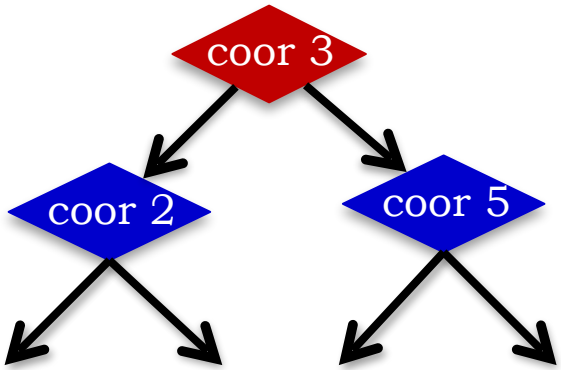
Question: Can we adapt to the dataset while keeping good performance on all (even adversarial) queries?



Bridging the Gap: Instance optimality?

Instance Optimality: best possible algorithm for the given dataset

- Focused goal:
  - "Best possible" in a class: here, tree-based space partition
  - Quality measure: here, Pr[success] of worst-case query
- [AB'21]: for Hamming space
  - Class = bit sampling/coordinate cut
  - Includes optimal worst-case LSH [IM'98]

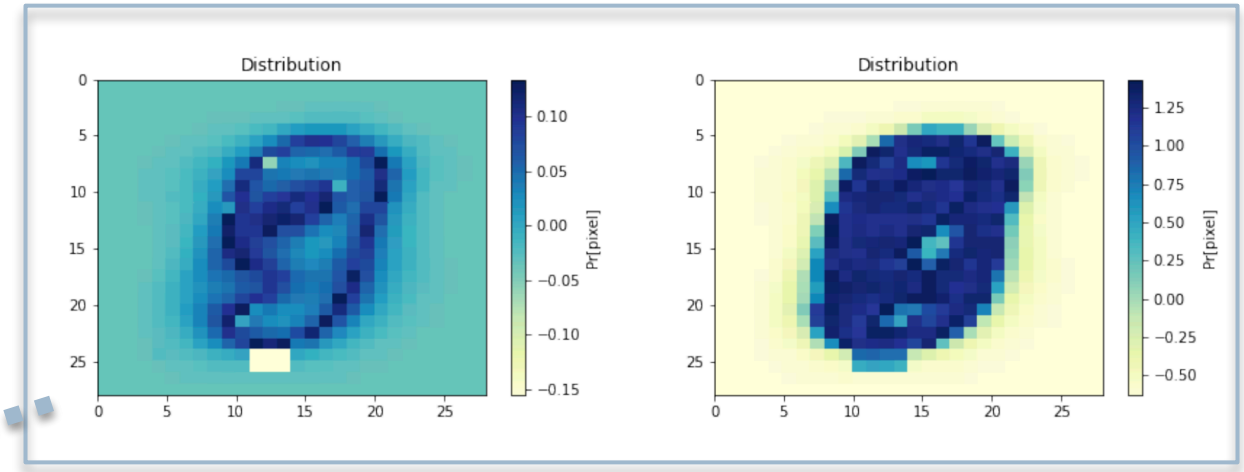


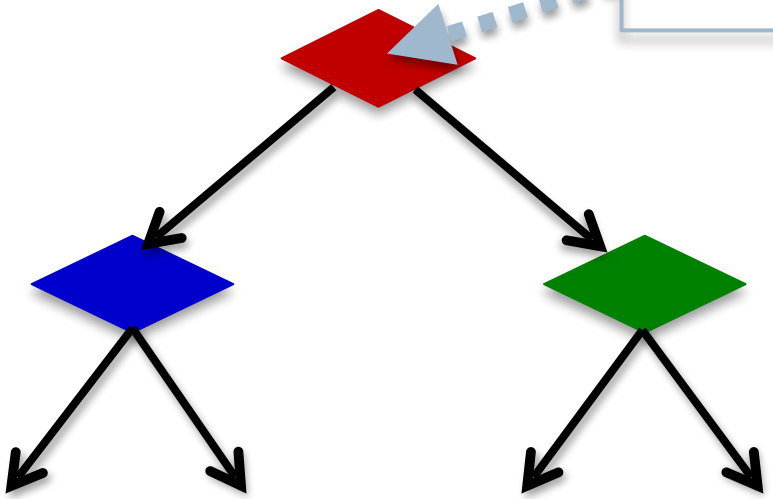
Our Results

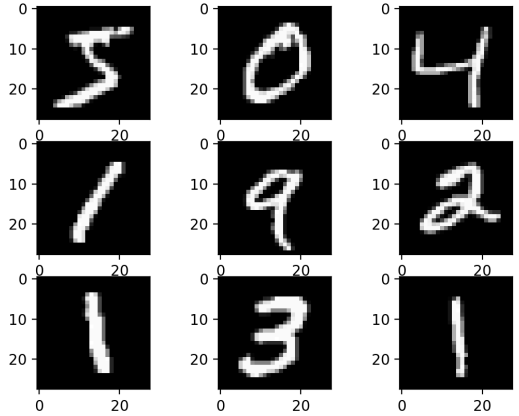
- We provide an algorithm for Hamming space with the following.
  - Can sample a data-optimized tree with:
    - $\Pr[success] \geq n^{-\rho}$ for every query, $\rho = 1/c$ (optimal LSH [IM'98])
  - Improvement over known algorithms with this guarantee:
    - Theoretically, we show for a mixture model (2 random clusters)
    - Empirically, we show improvement for MNIST and ImageNet
  - Bonus: in experiments, the *average* Pr[success] also improved

Optimized Distributions on MNIST

More weight in center of the images







Thank you!

Feel free to email:
Daniel Beaglehole
dbeaglehole@ucsd.edu