

LEARNING CONVEX BODIES IS HARD

(COLT 2009 – NAVIN GOYAL AND LUIS RADEMACHER)

SIGML S02E03

Purushottam Kar

Distribution Specific Learning

2

- Can we try to learn the concepts under certain “natural” distributions ?
- **[GR09]** : Convex bodies are hard to learn even under the uniform distribution
- More specifically, there are convex bodies which force every learning algorithm to draw at least $2^{\Omega(\sqrt{d/\epsilon})}$ samples from the uniform distribution
- **[KO'DS08]** Under the Gaussian distribution, learning is
 - ▣ possible in time $2^{\tilde{O}(\sqrt{d})}$
 - ▣ requires $2^{\Omega(\sqrt{d})}$ samples

Learning Convex Bodies

3

- Distance b/w bodies K and K' : $vol(K' \setminus K) / vol(K)$
- We say an algorithm has learnt a body K if it outputs a body K' such that $dist(K, K') < \varepsilon$
- Corresponds to the notion of error if $vol(K) = vol(K')$
- Plan of attack
 - ▣ Find a family of convex bodies that is pairwise dissimilar
 - ▣ Reduce the problem of learning within this family to the problem of learning convex bodies
 - ▣ Prove that learning within this family is hard
 - Prove that distinguishing one member of this family from any other requires exponentially many samples

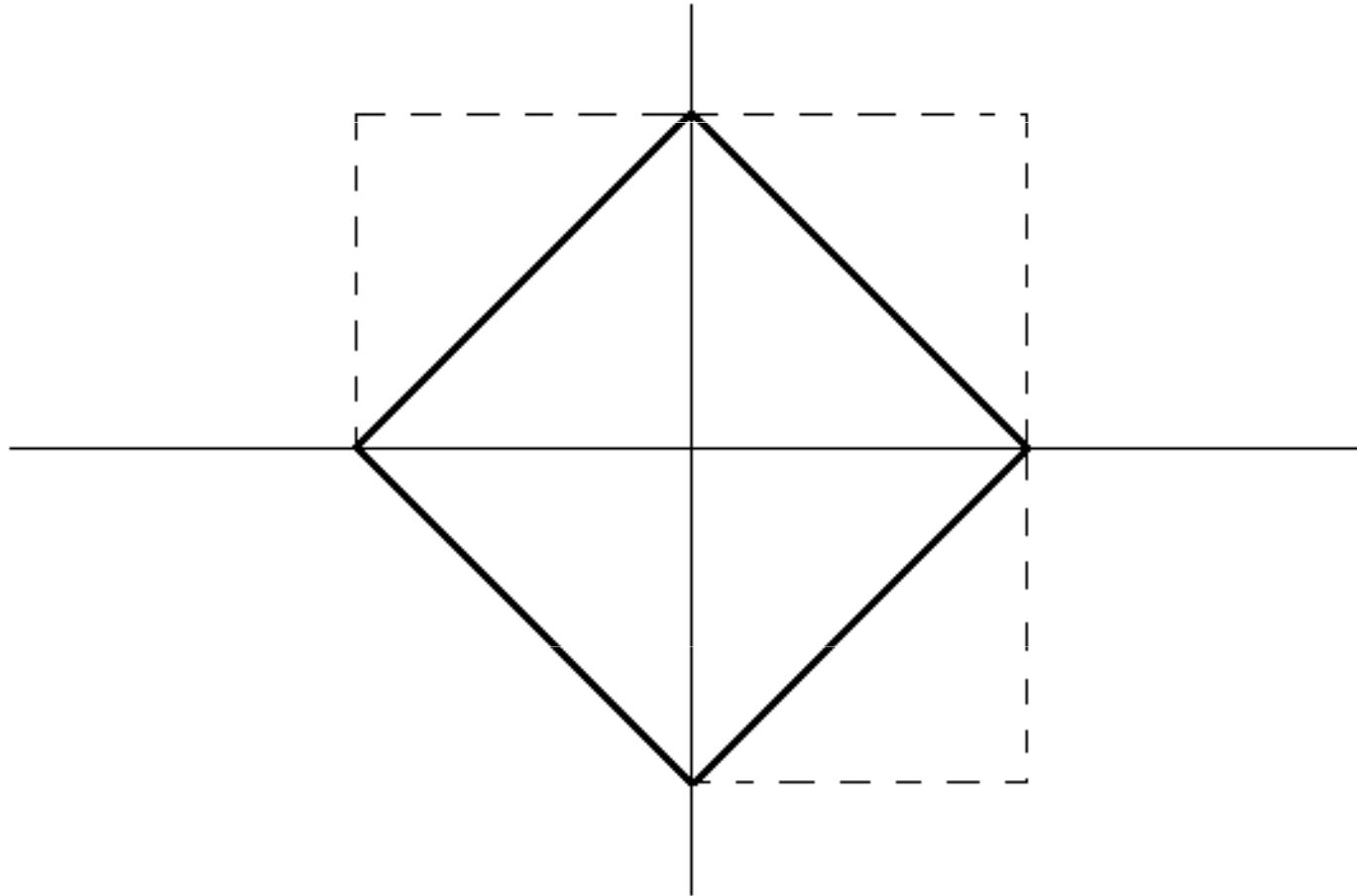
Error Correcting Codes

4

- A code on strings of length n over an alphabet Σ is a subset of Σ^n
- The distance parameter of a code is the minimum Hamming distance between any two of its codewords
- **[Gilbert-Varshamov]** For string length n over alphabets of size q , there exist codes of size $q^n / V_q(n, d-1)$ with distance parameter d
- For binary strings of length $2n$, there exist codes of size $2^n / V_2(n, d-1)$ with distance parameter $2d$ such that each codeword has weight n (i.e. non-zero at those many locations)

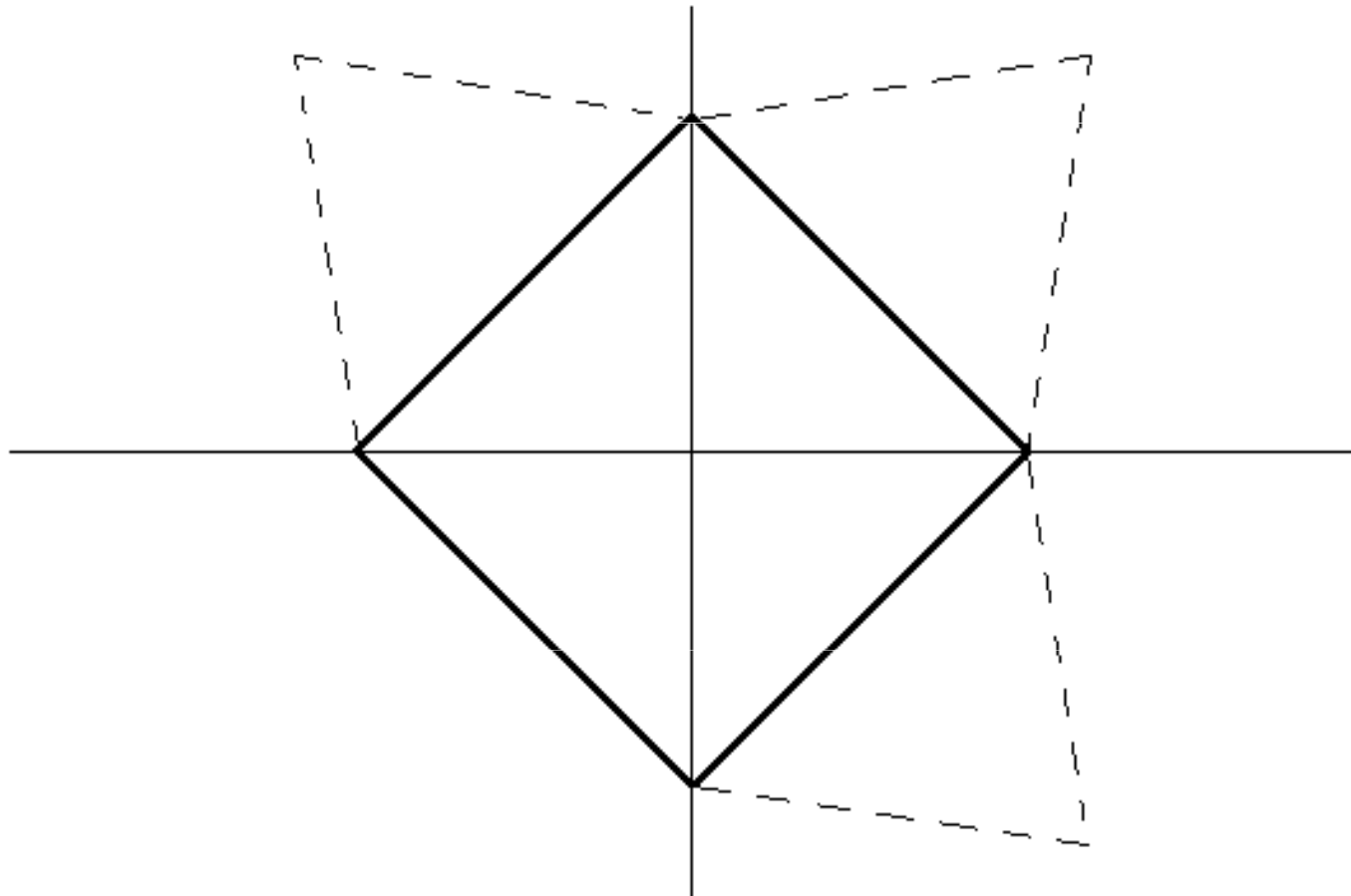
Cross Polytope with Peaks

5



No Mountain too high ...

6



The Inner Family

7

- Working in n dimensions – consider a code with weight $N/2$, minimum relative distance $1/4$ over binary strings of length N where $N = 2^n$
- There exist codes of size $2^{\Omega(n)}$ that satisfy these properties – choose such a code C
- Construct a family of polytopes – one corresponding to each codeword in the code
- Properties of the code ensure that any two polytopes in this family have a volume difference of $\frac{c_n}{4(n-1)}$
- Let the set \mathcal{I} index this family of polytopes

The Outer Family

8

- Choose a code C' over the alphabet \mathcal{I} over strings of length k with relative distance $1/2$
- For each codeword in this new code, construct a new product polytope, this time in $d = kn$ dimensions

$$(K_1, K_2, \dots, K_k) \rightarrow K_1 \times K_2 \times \dots \times K_k$$

- The size of this family of polytopes is $|\mathcal{I} / 4|^{k/2} = 2^{\Omega(kN)}$
- Any two polytopes in this family are at least distance $1 - e^{-k/(16n)} = 1 - e^{-d/(16n^2)}$ apart
- Choose n such that this distance is greater than 2ε
 - $n = 2^{\Omega(\sqrt{d/\varepsilon})}$ suffices

A further simplification

9

- Reduce the problem to one in which oracles are *discrete*
 - ▣ Random Oracle : returns a random region in the polytope instead of a point. Regions (either the diamond or one of the peaks) are returned with $\text{prob.} \propto \text{volume of the region}$
 - ▣ Membership Oracle : returns whether the query peak belongs to the polytope or not
- Thus there are only a finite number of deterministic algorithms that attempt to learn these bodies using at most q queries
 - ▣ think of each algorithm as a decision tree
 - ▣ there are at most $(N + 1)^{kq}$ leaves in any tree

Finishing off

10

- How often can any of these deterministic algorithms give the correct answer when learning a randomly chosen polytope ?
 - ▣ Note : there are more bodies than there are leaves
 - ▣ i.e. not more than $(\# \text{ leaves}) / (\# \text{ bodies in the family})$
 - ▣ this turns out to be at most $(N + 1)^{kq} / 2^{\Omega(kN)}$
 - ▣ This implies that even if the success probability desired is $1/2$, the number of queries needed is at least $2^{\Omega(\sqrt{d/\epsilon})}$
- Final Step : Apply *Yao's Minimax Principle* to argue for the same lower bound for randomized algorithms

Open Problems

11

- Under what conditions does the Surface area characterize efficient learnability ?
- Learning Convex Bodies with labeled samples under various distributions ?
 - ▣ Intuitively should be no more difficult than learning with positive samples alone
- Learning convex bodies with polynomially many facets in polynomial time ?
 - ▣ It is known that information theoretically, polynomially many samples are sufficient

References

12

- [GR09] Navin Goyal and Luis Rademacher, **Learning Convex Bodies is Hard**, COLT, 2009.
- [KO'DS08] Adam Klivans, Ryan O'Donnell and Rocco Servedio, **Learning Geometric Concepts via Gaussian Surface Area**, FOCS, 2008.