Learning mixtures of spherical Gaussians.

$$X = \sum_{i=1}^{k} w_i N(\mu_i, I) \quad \subset \mathbb{R}^n.$$

Given samples from X and you want to learn X.

(either: learn $\hat{X}$ s.t. $d_{TV}(X, \hat{X}) < \varepsilon$,
learn $\hat{\mu}_i$ s.t. (up to a perm.) $|\mu_i - \hat{\mu}_i| < \varepsilon$)

Note: Learning Non-spherical Gaussians has $n^{\Omega(k)}$ lower bounds in the SQ model.

Natural approach is method of moments.

[as long as $|M_i|$ not too large].

easy: approximate 1st d moments of X for any constant d.

Question: What are these moments?

$$X = D * G$$

discrete distr. over means.

$N(0, I)$

$$\mathbb{E}[X^{\otimes t}] = \sum_i \binom{t}{i} \mathrm{Sym}\left(\mathbb{E} D^{\otimes i} \otimes \underline{G^{\otimes t-i}}\right)$$

Not hard to deconvolve.

approximate $\mathbb{E}[D^{\otimes i}]$

Q Can we learn D from its low order moments?

A No. Consider 1-d. problem.

first ~~k~~ d moments. d+1 parameters

but X has 2k parameters.

$\Rightarrow$ Unless $d \geq 2k$. moments don't determine D.

(even up to small distance)

D, D' match moments,

N·D, N*D' also match moments.

embedd into n dimensions

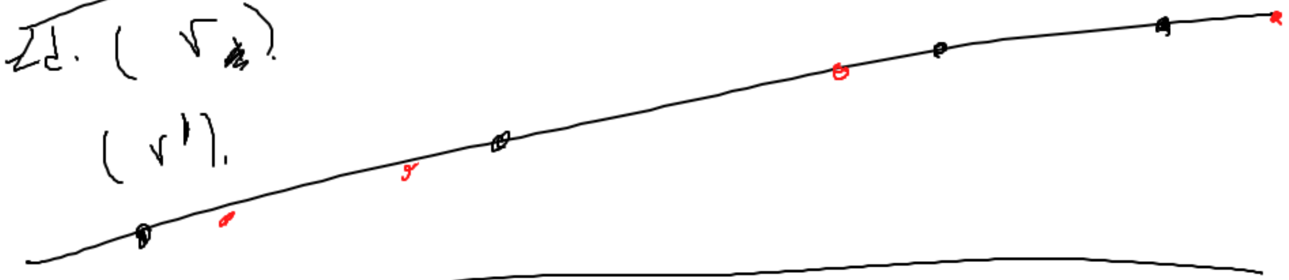
coef's of p = sym polys in roots

$x^d - \sigma_1 x^{d-1} + \dots \sigma_d \rightarrow \sum r_i^t$ for $t < d$

sym poly of deg d. (r_d)

" " (r^1).

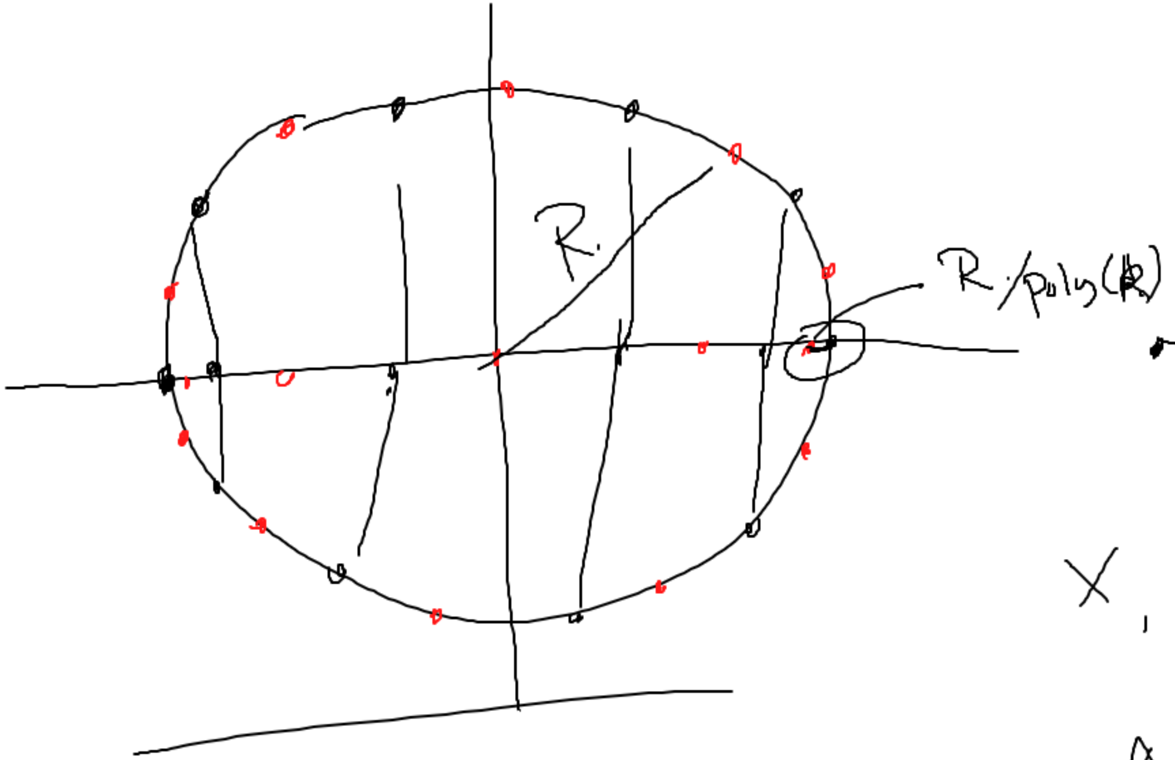
Sym — —

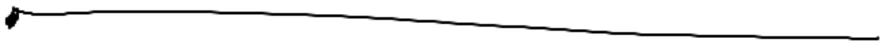


coef's of p' = — — —

Pick p, p' two degree d polys that differ by a constant. (w/ d real roots).

& let means of D be roots of p
means of D' be roots of p'



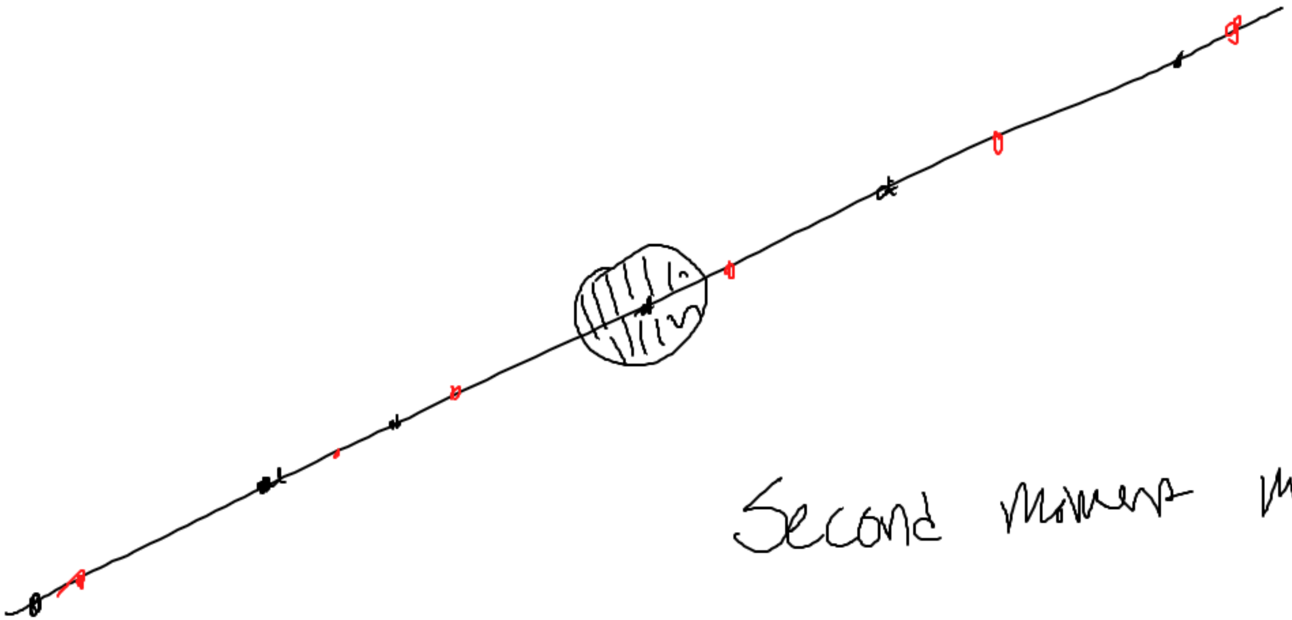


$X, \quad X'$

~~A+b~~ Close to distance 1 in d_{TV}.

any 1-d X approximate X by empirical distribution $* N(0,\Sigma)$

but first k moments match.

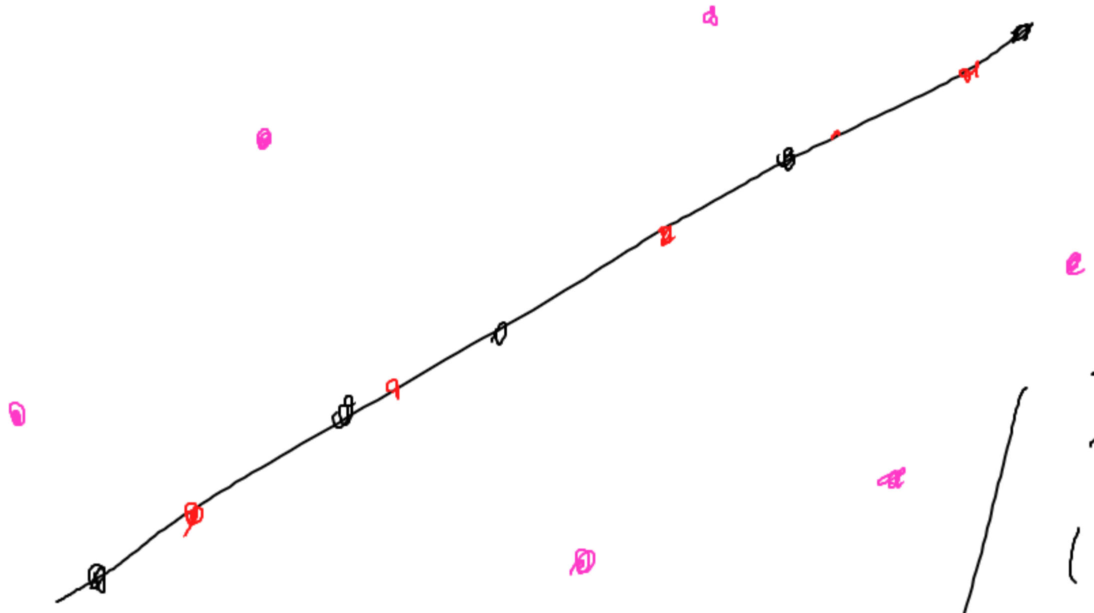


Second moment matrix of (X)

= I + Second mom of (D)

~~Span~~ rank $\le k$
Span $\subset$ span of μ_i's.

Suppose on a ~~Start~~ k
Subspace of dim.



reduce to
k-dim
problem.

to prove lower bounds
(say in SQ)

<u>want</u> is measures that
are spherically symmetric.

Impossible

$X = D * G.$

low deg moments of $X \longrightarrow$ low deg moments of D

$\longrightarrow \sum_{i=1}^{k} w_i\, q(\mu_i)$ for any low degree polys. q

$q(x) = p^2(x)$ for some p

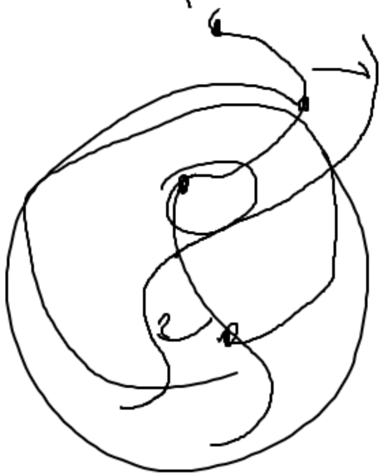
$= \sum \underline{w_i\, p^2(\mu_i)}$

$= 0$ iff $\boxed{p(\mu_i) = 0}$ for all i.

Are there low degree polys p s.t.

$p(u_i) = 0$ for all i?

Yes



Linear transformation

$$\underbrace{\{\text{deg} \le d \text{ polys}\}}_{\mathcal{P}} \xrightarrow{\binom{n+d}{d}} \underbrace{(P(u_1), P(u_2), \ldots, P(u_k))}_{\text{dim } k} \subset \mathbb{R}^k.$$

Kernel of $\mathcal{P}$.

Kernel has codim $\le k$.

$\boxed{\ \ }\ \mathbb{R}^{n'}$ $\binom{n'+d}{d} > k$ if $n' > d k^{1/d}$.

Compute moments of x → Compute D. → Compute space of low degree vanishing polynomials P_i (exactly)

V = Variety defined by these polynomials

= Set of pts that these polys vanish on.

$\mu_i \in V$. <u>intuition</u> V small

<u>Proof</u> $\dim($ ~~so~~ deg-d polys on V $) \leq$ Codim of space of defining polys.

$\leq R$.

$\rotatebox{90}{$\leq$}$

$\binom{\dim(V)+d}{d}$

$\Rightarrow$ <u>$\dim(V) \leq d R^{1/d}$.</u>

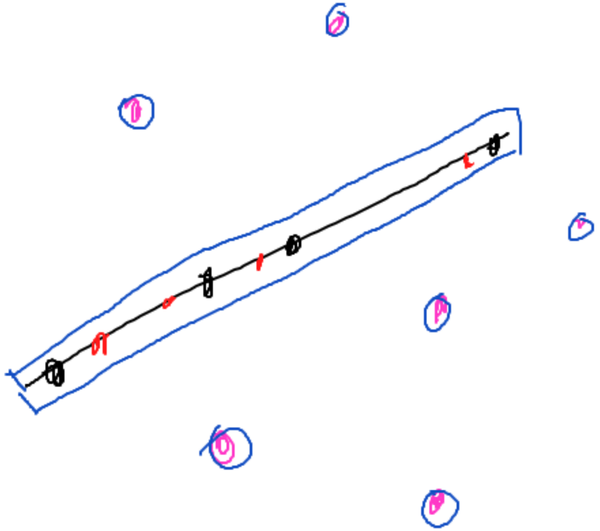
R

Idea "project" onto V.

do something exhaustive in random $\exp(\dim(V))$

Computing d dimensions takes n^d time.

$d = \lg(k)$.



reduce to $d\,k^{1/d} \approx d$. dimensional variety?

exhaustive then takes $\approx 2^d$ time.

quasi-poly alg.

Technical problems

1. V is low dimension but is it "simple"?

2. How do we project onto V?

3. How do we compute on V?

4. How do I deal with approx. defining polynomials?

Idea ~~Ep~~ replace V w/ a point cloud.

~~Set~~ Vector space $\mathcal{U}$ of degree d polys $\mathbb{R}$

$\operatorname{Codim}(\mathcal{U}) \le k$.

we'll get back to [illegible]

$$S = \{ x \in \mathbb{R}^n : |x| < R, \ |p(x)| < \underline{\delta |p|} \text{ for all } p \in \mathcal{U} \}$$

<u>Want</u> for $\varepsilon > 0$ to find a small ε-cover of S.

~~Thm~~

<u>Thrm</u> If $\varepsilon > \delta^{\frac{1}{2d}} \operatorname{poly}(R\, n\, d\, k)$ then there exists an ε-cover of size at most

$$(2(R/\varepsilon) \cdot d k n)^{O(d^2 k^{1/d})}$$

expect cover size should be $\approx$

$O(R/\varepsilon)^{\dim(Y)}$

$dk^{1/d}$ we get $d^2k^{1/d}$

$$\boxed{\varepsilon > \delta^{1/2d}(\cdots).}$$

Technicalities

(1) $\quad U \subset$ hom. degree-d polynomials.

(2) $\quad \delta < (\varepsilon / R d n)^{c \cdot d}$. $\qquad P(x) = A \cdot x^{\otimes d}$. $\qquad |P| := |A|_2$.

$|p|$ is spherically symmetric.

Define $\quad f(\underline{\varepsilon}, \underline{R}, \underline{d}, \underline{k}, \underline{n}, \underline{\delta}) =$ ~~[illegible]~~ worst case cover size.

Prove a recursive upper bound on f.

$$\mathbb{R}^n = \mathbb{R}^{n'} \times \mathbb{R}^{n-n'}$$

$$(\underset{\downarrow}{x} \quad , \quad \underset{\downarrow}{y})$$

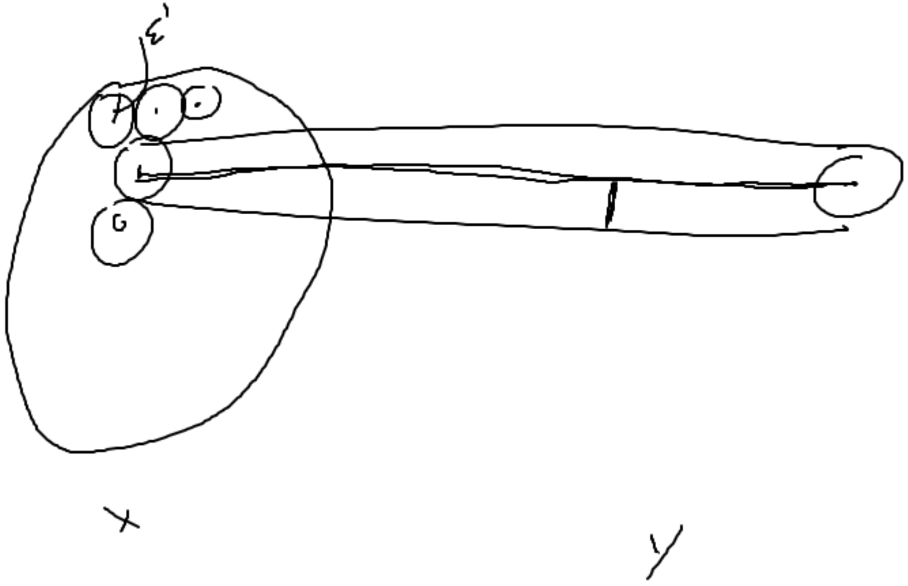
n' as smallish

$W = \{$ polys in U that are degree 1 in x & degree $d-1$ in $y\}$

$\mathrm{codim}(W) \leq k$.

if U is codim k in V.

then $U \cap W$ is codim $\leq k$ in $V \cap W$.



ε' very small

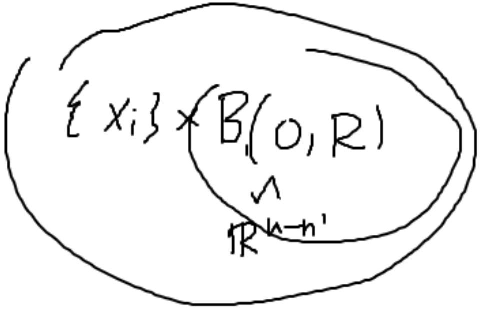
small enough

If I change δ by a bit....

It is enough to cover

this center line. $\{x_0\} \times \mathbb{R}^{n-n'}$

$\{X_i\}$



$A_{x_i} : W \longrightarrow$ { hom deg $d-1$ polys in y }

$P(x, y) \longrightarrow P(x_i, y)$

$\cup$
deg 1, deg $d-1$
polys in U.

S_{x_i} = { non vanishing pols w/ $x = x_i$ }

η not too small

$n^1 \times (\cdots)$.

If $q = A_{x_i} P$. $|q| > \eta |p|$.

(x_i, y)

Then S_{x_i} must nearly vanish on q.

nearly vanishing

U_{x_i} = span of the singular ~~vectors~~ vectors of A_{x_i} w/
singular value $\geq \eta$.

$\delta|p| = (\delta/\eta)|q|$.

$S_{x_i} \subset$ { pts that nearly vanish on U_{x_i} }

recursive version of original problem.

Say x_i is good if $\operatorname{Codim}(U_{x_i}) \le k'$

If x_i is good then the cylinder can be covered by a set of size.

$$f(d-1, k', n-n', R, \tilde{\varepsilon}, \tilde{\delta}/\eta)$$

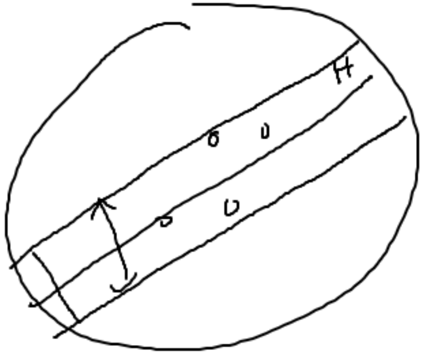
What about the not good pts

$$\delta \ \cancel{\ll} \ \varepsilon^{d} (\ldots) \qquad \eta \approx \mathrm{poly}(\varepsilon)$$

Then I can cover all of the cylinders over all of the good points w/ an appropriate # of cov[er]s.

$$f(\ldots) \le (\ldots)^{d^2 k^{1/d}} \qquad k' \approx k^{\frac{d-1}{d}}$$

Prop There exists a subspace $H < \mathbb{R}^{n'}$ of dimension $\leq 2k/k'$ s.t. all of the bad pts are close to H.



Cover $\underbrace{H \times \mathbb{R}^{n-n'}}$ as long as $n' \gg k/k' \gtrsim k^{1/d}$.

subspace of dim $\mathbb{R}^{n-n'+2k/k'}$ Then I should be fine.

$+ f(d, k, R, \tilde{\varepsilon}, \delta, \underline{n-n'+2k/k'})$