نظریه اطلاعات، آمار و یادگیری

دکتر یاسایی

دانشگاه صنعتی شریف

مهندسی برق

برنا خدابنده ۴۰۰۱۰۹۸۹۸

تمرین شماره Final - Q1

تاریخ: 1403/03/05

Problem 1) Variational form χ^2: $\sup_{h:\mathcal{X}\to\mathbb{R}} \left\{ \mathbb{E}_P[h(x)] - \mathbb{E}_Q\left[\frac{h^2(x)}{4}\right] - 1 \right\} = \chi^2(P\|Q)$

Since: $D_f(P\|Q) = \sup_h \left\{ \mathbb{E}_P[h(x)] - \mathbb{E}_Q[f^*(h(x))] \right\}$, $\chi^2: f(x) = x^2 - 1 \rightsquigarrow f^*(y) = \frac{1}{4}y^2 + 1$

let: $h(x) = f(x) + c$, $c \in \mathbb{R}$, $f: \mathcal{X}\to\mathbb{R} \rightsquigarrow \sup_{h:\mathcal{X}\to\mathbb{R}} \{\dots\} = \sup_{f:\mathcal{X}\to\mathbb{R}} \sup_{c\in\mathbb{R}} \{\dots\}$

$$\chi^2(P\|Q) = \sup_{h:\mathcal{X}\to\mathbb{R}} \{\dots\} = \sup_{f:\mathcal{X}\to\mathbb{R}} \sup_{c\in\mathbb{R}} \left\{ \mathbb{E}_P[f(x)+c] - \mathbb{E}_Q\left[\frac{1}{4}(f(x)+c)^2\right] - 1 \right\}$$

$$= \sup_{f:\mathcal{X}\to\mathbb{R}} \sup_{c\in\mathbb{R}} \{ \underbrace{\mathbb{E}_P[f(x)] + c - \frac{1}{4}\mathbb{E}_Q[f^2(x)] - \frac{1}{2}\mathbb{E}_Q[f(x)]c - \frac{1}{4}c^2 - 1}_{\mathcal{L}(x;c,f)} \}$$

@ $c_0 = \arg\max \mathcal{L}(x;c,f)$: $\frac{\partial \mathcal{L}(x;c,f)}{\partial c} = 0 \rightsquigarrow 1 - \frac{1}{2}\mathbb{E}_Q[f(x)] - \frac{1}{2}c_0 = 0$

$\Rightarrow c_0 = 2 - \mathbb{E}_Q[f(x)] \rightsquigarrow \chi^2 = \sup_{f:\mathcal{X}\to\mathbb{R}} \mathcal{L}(x;c_0,f)$

$$\chi^2(P\|Q) = \sup_{f:\mathcal{X}\to\mathbb{R}} \{ \mathbb{E}_P[f(x)] + 2 - \mathbb{E}_Q[f(x)] - \frac{1}{4}\mathbb{E}_Q[f^2(x)] - 1 - \mathbb{E}_Q[f(x)] + \frac{1}{2}\mathbb{E}_Q[f(x)]^2 \dots$$

$$\underbrace{-\frac{1}{4}(2 - \mathbb{E}_Q[f(x)])^2}_{-1+\mathbb{E}_Q[f(x)] - \frac{1}{4}\mathbb{E}_Q[f(x)]^2} \} = \sup_{f:\mathcal{X}\to\mathbb{R}} \left\{ \mathbb{E}_P[f(x)] - \mathbb{E}_Q[f(x)] + \frac{1}{4}\left(\mathbb{E}_Q[f(x)]^2 - \mathbb{E}_Q[f(x)^2]\right) \right\}$$

$$\Rightarrow \boxed{\chi^2(P\|Q) = \sup_{f:\mathcal{X}\to\mathbb{R}} \left\{ \mathbb{E}_P[f(x)] - \mathbb{E}_Q[f(x)] - \frac{1}{4}\mathrm{Var}_Q[f(x)] \right\}}$$

$$\chi^2(P_{XY}\|Q_XQ_Y) = \sup_{f:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}} \left\{ \mathbb{E}_P[f(x,y)] - \mathbb{E}_Q[f(x,y)] - \frac{1}{4}\mathrm{Var}_Q[f(x,y)] \right\}$$

$\to Q_{XY} = Q_X Q_Y$
X, Y are independent

$$\geq \sup_{\substack{f(x,y)=h(x)+g(y)\\ h:\mathcal{X}\to\mathbb{R}\\ g:\mathcal{Y}\to\mathbb{R}}} \{\dots\} = \sup_{g:\mathcal{Y}\to\mathbb{R}} \sup_{h:\mathcal{X}\to\mathbb{R}} \left\{ \mathbb{E}_{P_{XY}}[h(x)+g(y)] - \mathbb{E}_{Q_XQ_Y}[h(x)+g(y)] - \frac{1}{4}\underbrace{\mathrm{Var}_{Q_XQ_Y}[h(x)+g(y)]}_{\text{by independence}} \right\}$$

Sum of variances

$$= \sup_{g:\mathcal{Y}\to\mathbb{R}} \sup_{h:\mathcal{X}\to\mathbb{R}} \left\{ \mathbb{E}_{\cancel{P_{XY}}\,P_X}[h(x)] + \mathbb{E}_{\cancel{P_{XY}}\,P_Y}[g(y)] - \mathbb{E}_{\cancel{Q_XQ_Y}\,Q_X}[h(x)] - \mathbb{E}_{\cancel{Q_XQ_Y}\,Q_Y}[g(y)] - \frac{1}{4}\left(\mathrm{Var}_{Q_X}(h(x)) + \mathrm{Var}_{Q_Y}(g(y))\right) \right\}$$

$$= \sup_{g:\mathcal{Y}\to\mathbb{R}} \sup_{h:\mathcal{X}\to\mathbb{R}} \left\{ \underbrace{\left(\mathbb{E}_{P_X}[h(x)] - \mathbb{E}_{Q_X}[h(x)] - \frac{1}{4}\mathrm{Var}_{Q_X}[h(x)]\right)}_{\chi^2(P_X\|Q_X)} + \underbrace{\left(\mathbb{E}_{P_Y}[g(y)] - \mathbb{E}_{Q_Y}[g(y)] - \frac{1}{4}\mathrm{Var}_{Q_Y}[g(y)]\right)}_{\chi^2(P_Y\|Q_Y)} \right\}$$

$$\Rightarrow \boxed{\chi^2(P_{XY}\|Q_XQ_Y) \geq \chi^2(P_X\|Q_X) + \chi^2(P_Y\|Q_Y)}$$

نظریه اطلاعات، آمار و یادگیری

دکتر یاسایی

دانشگاه صنعتی شریف

مهندسی برق

برنا خدابنده ۴۰۰۱۰۹۸۹۸

تمرین شماره ..Final_Q2

تاریخ: 1403/03/06

Problem 2) $y = Ax + z$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times d}$, $z \sim N(0, \sigma^2 I_n)$)

k-sparse : $x \in \mathbb{R}^d$, $\|x\|_0 \leq k$, assuming $k \leq \frac{d}{2}$ (sparsity)

minimax bounds : $E := \min_{\hat{x}(y)} \max_{x \in S_k^d} [\|\hat{x} - x\|^2]$, S_k^d : d dimentional k sparse

phase 1 : $h(x) = \mathbb{E}[-\log f_x(x)]$, $P_y = N(0, \beta I)$

$$D(P_x \| P_y) = \mathbb{E}_{x\sim P_x}\left[\log\left(\frac{dP_x(x)}{dP_y(x)}\right)\right] = \mathbb{E}_{x\sim P_x}\left[\log\left(\frac{f_x(x)}{f_y(x)}\right)\right] = -\mathbb{E}_{x\sim P_x}[\log(f_y(x))] - h(x)$$

reminder : $Z \sim N(\mu, \Sigma) \Rightarrow f_z(z) = (2\pi)^{-d/2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$

$$Y : \Sigma = \beta I,\ \mu = 0 \Rightarrow \det(\Sigma) = \beta^\ell \Rightarrow f_y(x) = (2\pi\beta)^{-\ell/2} \exp\left(-\frac{1}{2\beta}\|x\|_2^2\right)$$

$$\mathbb{E}_{x\sim P_x}[\log(f_y(x))] = \mathbb{E}_{x\sim P_x}\left[\log\left((2\pi\beta)^{-\ell/2}\exp\left(-\frac{1}{2\beta}\|x\|_2^2\right)\right)\right] = -\frac{\ell}{2}\log(2\pi\beta) - \frac{\log e}{2\beta}\mathbb{E}_{x\sim P_x}[\|x\|^2]$$

$$\Rightarrow \boxed{D(P_x \| P_y) = \frac{\ell}{2}\log(2\pi\beta) + \frac{\log e}{2\beta}\mathbb{E}[\|x\|_2^2] - h(x)}$$

$$\Rightarrow h(x) = \frac{\ell}{2}\log(2\pi\beta) + \frac{\log e}{2\beta}\mathbb{E}[\|x\|_2^2] - D(P_x \| P_y)$$

let : $\beta = \frac{t}{\ell} \Rightarrow \mathbb{E}_{x\sim P_y}[\|x\|_2^2] = \sum \mathbb{E}_{x\sim P_y}[x_i^2] = \ell \cdot \frac{t}{\ell} = t$

$$h(x) = \frac{\ell}{2}\log\left(\frac{2\pi t}{\ell}\right) + \frac{\ell}{2t}\log(e)\underbrace{\mathbb{E}[\|x\|_2^2]}_{\text{increase this}} - \underbrace{D(P_x \| P_y)}_{\text{decrease this}}$$

$\boxed{P_x = P_y \text{ is optimal}}$

maximal

$D(P_x \| P_y) \geq 0 \Rightarrow$ best scenario $\underbrace{P_x = P_y}_{\text{minimal } D(P_x\|P_y)} \Rightarrow X \sim N(0, \frac{t}{\ell} I) \rightarrow \mathbb{E}[\|x\|_2^2] = t$

$\Rightarrow \boxed{h(x) : \text{maximized at } X \sim N(0, \frac{t}{\ell} I)}$, $h(x) = \frac{\ell}{2}\log\left(\frac{2\pi t}{\ell}\right) + \frac{\ell}{2}\log(e)$

$$\Rightarrow \boxed{h(x) = \frac{\ell}{2}\log\left(\frac{2\pi t e}{\ell}\right)}$$

$$I(x;y) = \mathbb{E}_{y\sim P_y}\left[\int_x f_{x|y}(x|y)\log\left(\frac{f_{x|y}(x|y)}{f_x(x)}\right)dx\right]$$
$$h(x|y) = \mathbb{E}_{y\sim P_y}\left[-\int_x f_{x|y}(x|y)\log(f_{x|y}(x|y))dx\right]$$
$$\Rightarrow I(x;y) + h(x|y) = \mathbb{E}_{y\sim P_y}\left[-\int_x f_{x|y}(x|y)\log(f_x(x))dx\right]$$

$$\Rightarrow I(x;y) + h(x|y) = -\int_x \underbrace{\mathbb{E}_{y\sim P_y}[f_{x|y}(x|y)]}_{f_x(x)}\log(f_x(x))dx = -\int_x f_x(x)\log(f_x(x))dx = h(x)$$

$$\Rightarrow \boxed{I(x;y) = h(x) - h(x|y) = I(x;y) = I(y;x) = h(y) - h(y|x)}$$

similarly, changing x, y

in this case:

$$h(y|x) = h(Ax+z|x) \rightsquigarrow y = Ax+z \rightsquigarrow f_{y|x}(y|x) = f_z(y-Ax)$$

$$h(y|x) = \mathbb{E}_{x\sim P_x}\left[-\int_y f_{y|x}(y|x)\log(f_{y|x}(y|x))dy\right] = \mathbb{E}_{x\sim P_x}\Big[\underbrace{-\int_y f_z(y-Ax)\log(f_z(y-Ax))dy}_{h(z),\ z=y-Ax:\ dz=dy}\Big]$$

change of integral variable

$$\Rightarrow \boxed{h(y|x) = h(z)}$$: obvious since only ambiguity is from z

$$\Rightarrow \boxed{y = Ax+z \ : \ I(x;y) = h(y) - h(z)}$$

Phase 2:

$\mathcal{B}$: signed permutations

notation: $\mathcal{B} = [b_{ij}],\ b_{ij} \in \{-1,0,1\}$, only one nonzero entry per column/row

let: $\forall i \neq i_k : b_{ki} = 0,\ b_{ka_k} \neq 0 \Rightarrow \forall i: b_{ki} = \delta_{ia_k} S_k,\ S_k = \mathrm{Sign}(b_{ki_k})$

$$(BB')_{ij} = \sum_{k=1}^{n} B_{ik}B'_{kj} = \sum_{k=1}^{n} \delta_{ka_i}\delta_{ja_k} S_i S'_k = \underbrace{S_i S'_{a_i}}_{S''_i}\delta_{j\underbrace{a'_{a_i}}_{a''_i}} = \boxed{S''_i\,\delta_{ja''_i} \rightsquigarrow \in \mathcal{B}}$$

Another interpretation: $\forall B \in \mathcal{B} : \exists S_l, P, S_r : P \in$ permutation matrix & $B = S_l P S_r$

S : diagonal with ± 1 entries

obvious since we can just take the sign of each column/row to S_l, S_r

$$\boxed{BB' = S_e P P' S'_r = S_e P'' S'_r = S_e B''_r = B'' \in \mathcal{B}}$$

the last equality holds since we are just changing signs.

it is well know that the product of two permutations is a permutation

$$\boxed{\forall B \in \mathcal{B} : B = S_\ell P = P S_r,\ P_{ij} = |b_{ij}|,\ (S_r)_{ii} = \mathrm{sign}(b_{a_i i}),\ (S_\ell)_{ii} = \mathrm{sign}(b_{i a_i})}$$

if: $x \in S_k^\ell$

D: diagonal, nonzero diags: $Dx \in S_k^\ell$: obvious since $(Dx)_i = D_{ii}x_i$ $\neq 0$

P: permutation $\Rightarrow$ $Px \in S_k^\ell$: obvious since only indices change

$$\rightarrow Bx = P\underbrace{Sx}_{y \in S_k^\ell} = Py \in S_k^\ell \Rightarrow \boxed{\forall x \in S_k^\ell : Bx \in S_k^\ell}$$

let: $B = PS \Rightarrow B^T = S^T P^T = SP^T \Rightarrow BB^T = PSSP^T$

$(SS)_{ii} = S_{ii}^2 = 1 \Rightarrow SS = I \Rightarrow BB^T = PP^T$, P: permutation

$$\rightarrow PP^T = I \Rightarrow \boxed{BB^T = I}$$

also obvious: $(BB^T)_{ij} = \sum_{k=1}^{n} B_{ik}B_{jk} = \sum s_i \delta_{k a_i} s_j \delta_{k a_j} = s_i s_j \delta_{a_i a_j} = s_i^2 \delta_{ij} = \delta_{ij}$

(with δ_{ij} written above $\delta_{a_i a_j}$)

$$\rightarrow \boxed{(BB^T)_{ij} = \delta_{ij} \Rightarrow BB^T = I}$$

$$\rightarrow \|Bx - Bx'\|_2^2 = \|B(x-x')\|_2^2 = (B(x-x'))^T(B(x-x')) = (x-x')^T \underbrace{B^T B}_{I} (x-x)$$

$B \in \mathcal{B} \Rightarrow B^T \in \mathcal{B} \Rightarrow (B^T)(B^T)^T = B^T B = I \Rightarrow \|Bx - Bx'\|_2^2 = \|x - x'\|_2^2$

$$\Rightarrow \boxed{\|Bx - Bx'\|_2 = \|x - x'\|_2}$$

two ways to view this:

one:

$B = SP$, $S: 2^\ell$ possibilities, $P: \ell!$ permutations $\Rightarrow \boxed{|B| = 2^\ell \ell!}$

two:

1's column: $\ell \times 2$, 2'nd column: $2\times(\ell-1)$, 3'rd: $2\times(\ell-2)$, ...

choices, sign, sign, remaining choices

$\Rightarrow |B| = (2\ell)(2(\ell-1))(\dots)(2\times2)(2\times1) = \boxed{2^\ell \times \ell! = |B|}$

$\tilde{B} = B_0 B$, $B_0, B \in \mathcal{B}$, $BB^T = I \Rightarrow \forall X \in \mathcal{B}: X$ is invertible

$\tilde{B}_1 = B_0 B_1$, $\tilde{B}_2 = B_0 B_2 \Rightarrow \tilde{B}_1 - \tilde{B}_2 = B_0(B_1 - B_2)$, B_0: full rank

$\Rightarrow \tilde{B}_1 = \tilde{B}_2$ iff $B_1 = B_2 \Rightarrow \forall B \in \mathcal{B}: \tilde{B} = B_0 B$ is unique,

$\tilde{B} \in \mathcal{B}$, from the pigeonhole principle, if we choose all $B \in \mathcal{B}$

we will generate all of $\mathcal{B}$ using $B_0 B$., unique $\Rightarrow$ uniform

$\Rightarrow \boxed{\mathbb{P}[b = B] = \mathbb{P}[B_0 b = B_0 B] = \mathbb{P}[\tilde{b} = \tilde{B}] = \frac{1}{|\mathcal{B}|} \Rightarrow \text{uniform}}$

$\mathbb{E}[Bx] = \mathbb{E}[B]x$, $\mathbb{E}[B_{ij}] = \mathbb{P}[B_{ij} = 1] - \mathbb{P}[B_{ij} = -1] \overset{\text{uniform}}{=} 0$

$\Rightarrow \boxed{\mathbb{E}[Bx] = \mathbb{E}[B]x = 0}$

let: $\mathbb{E}[Bxx^T B] = M \Rightarrow \forall B_0 \in \mathcal{B}: B_0 M B_0^T = B_0 \mathbb{E}_{B\sim \mathrm{Unif}(\mathcal{B})}[Bxx^T B^T] B_0^T$

$= \mathbb{E}_{B\sim\mathrm{Unif}(\mathcal{B})}[B_0 B x x^T (B_0 B)^T] = \mathbb{E}_{\tilde{B}\sim\mathrm{Unif}(\mathcal{B})}[\tilde{B} x x^T \tilde{B}^T] = M = B_0 M B_0^T$, $B_0 B_0^T = I$

$\Rightarrow MB_0 = B_0 M \quad \forall B_0 \in \mathcal{B}$

$B_1 = B^{+ij}$: some $B \in \mathcal{B}$ such that $B_{ij} = 1$

$B_2 = B^{-ij}$: same as B^{+ij}, except $B_{ij} = -1$ $\rightsquigarrow$ obviously $B^{-ij} \in \mathcal{B}$

$$\Rightarrow \left.\begin{matrix} MB^{+ij} = B^{+ij}M \\ MB^{-ij} = B^{-ij}M \end{matrix}\right\} \Rightarrow M\underbrace{(B^{+ij} - B^{-ij})}_{2\Delta^{ij}} = \underbrace{(B^{+ij} - B^{-ij})}_{2\Delta^{ij}}M \Rightarrow \Delta^{ij}M = M\Delta^{ij}$$

$$(\Delta^{ij}M)_{\ell m} = \sum_{k=1}^{n} \Delta^{ij}_{\ell k} M_{km} = \sum_{k=1}^{n} \delta_{\ell i}\delta_{kj} M_{km} = \delta_{\ell i} M_{jm} = (M\Delta^{ij})_{\ell m} = \sum_{k=1}^{n} M_{\ell k}\Delta^{ij}_{km}$$

$$= M_{\ell i}\delta_{jm} = M_{jm}\delta_{\ell i} \rightsquigarrow \left.\begin{matrix} M_{jm} = 0 \quad \forall j \neq m \rightsquigarrow \text{diagonal} \\ \hookrightarrow M_{jj} = M_{\ell\ell} \quad \forall j, \ell \end{matrix}\right\} \Rightarrow \boxed{M = \alpha I}$$

$$M = \alpha I = \mathbb{E}[Bxx^T B^T] \rightsquigarrow tr(\alpha I) = \ell\alpha = tr(\mathbb{E}[Bxx^T B]) = \mathbb{E}[tr(\ldots)]$$

$$= \mathbb{E}[tr(x^T B^T B x)] = \mathbb{E}[tr(x^T I x)] = tr(x^T x) = \|x\|^2 = \ell\alpha \Rightarrow \alpha = \frac{\|x\|^2}{\ell}$$

$$\Rightarrow \boxed{\mathbb{E}[Bxx^T B^T] = \frac{\|x\|_2^2}{\ell} I}$$

phase 3: $A := \{x = \frac{1}{\sqrt{k}}(x_1, \ldots, x_\ell) : x_i \in \{-1, 0, 1\}, \|x\|_0 = k\}$

$D(A, \|\cdot\|_2, \frac{1}{2})$:

$\forall x \in A : \|x\|_2^2 = \frac{\sum_{i=1}^{\ell} x_i^2}{k} = \frac{k}{k} = 1$ $\rightsquigarrow$ points lie on a sphere

let: $x_0 = \frac{1}{\sqrt{k}}(\overbrace{1, 1, \ldots, 1}^{k}, 0, 0, \ldots, 0)$, $\forall x \in A \; \exists B \in \mathcal{B} : x = Bx_0$

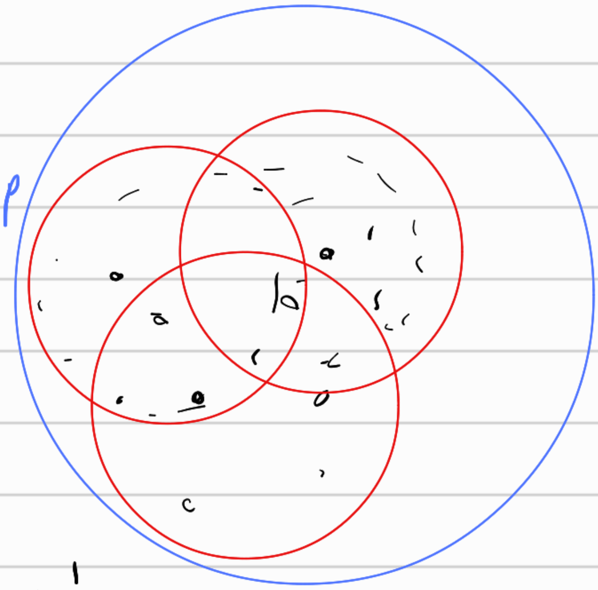
in this case, only the first k columns of B matter, thus $|A| < |B|$

by symmetry, if we choose B uniformly from $\mathcal{B}$, we get uniform on x.

$$\mathbb{P}[X=x] = \mathbb{P}[Bx_0 = x] = \mathbb{P}[B \in \{B \mid \text{first } k \text{ columns}\}] = \frac{(\ell-k)! \times 2^{\ell k} \times k!}{2^\ell \cancel{\ell!}}$$

$$\Rightarrow \boxed{\forall x: \mathbb{P}[Bx_0 = x] = \left[2^k \binom{\ell}{k}\right]^{-1} \rightsquigarrow \text{uniform on } A,\ |A| = 2^k \binom{\ell}{k}}$$

$N_{1/2}$ حجم شکل با فاصله کمتر از $\frac{1}{2}$ از ستاره اطراف



$$\sum_{i=1}^{P} N_{1/2}(p) \geqslant |A| \rightsquigarrow \text{if we find are independent of } p$$

$$\underbrace{P(A, \|\cdot\|, \tfrac{1}{2}) \times |N_{1/2}|}_{\text{ignoring overlaps.}} \geqslant |A| \Rightarrow P(A, \|\cdot\|, \tfrac{1}{2}) \geqslant \frac{|A|}{|N_{1/2}|}$$

$$A = \left\{x = \left\{0, \tfrac{1}{\sqrt{k}}, \tfrac{-1}{\sqrt{k}}\right\}^n : \|x\|_0 = k\right\}$$

$$\|x - x_0\|_2^2 = \sum (x_i - x_{0i})^2 \ ; \ (x_i - x_{0i})^2 : \begin{cases} (0, \frac{\pm 1}{\sqrt{k}}), (\frac{\pm 1}{\sqrt{k}}, 0) : \frac{1}{k} \\ (0,0), (\pm\frac{1}{\sqrt{k}}, \pm\frac{1}{\sqrt{k}}) : 0 \\ (\pm\frac{1}{\sqrt{k}}, \mp\frac{1}{\sqrt{k}}) : \frac{4}{k} \end{cases} \Rightarrow \geqslant \frac{1}{k} \mathbb{1}\{x_{0i} \neq x_i\}$$

$$\Rightarrow \|x - x_0\|_2^2 \geqslant \frac{1}{k}\|x - x_0\|_0 \rightsquigarrow N_{1/2} = \{x' \in A : \|x' - x\|_2^2 \leqslant \tfrac{1}{2}\} \subseteq \{x' \in A : \tfrac{1}{k}|x' - x|_0 \leqslant \tfrac{1}{2}\}$$

$$= \{x' \in A : |x' - x|_0 \leqslant \tfrac{k}{2}\} \Rightarrow |N_{1/2}| \leqslant |\{x' \in A : |x - x'|_0 \leqslant \tfrac{k}{2}\}| = \sum_{i=0}^{k/2} \binom{\ell}{i} 2^i$$

$$\rightsquigarrow |N_{1/2}| \leqslant \sum_{i=0}^{k/2} \binom{\ell}{i} 2^i \leqslant \binom{\ell}{k/2} 3^{k/2} : \text{choose } \tfrac{k}{2} \text{ indices, change to something, at most } \tfrac{k}{2} \text{ different}$$

$$\frac{|N_{1/2}|}{|A|} \leqslant \frac{\binom{\ell}{k/2} 3^{k/2}}{\binom{\ell}{k} 2^k} = \left(\frac{3}{4}\right)^{k/2} \frac{\binom{\ell}{k/2}}{\binom{\ell}{k}} = \left(\frac{3}{4}\right)^{k/2} \cdot \frac{k!(\ell-k)!}{(\frac{k}{2})!(\ell - \frac{k}{2})!} = \left(\frac{3}{4}\right)^{k/2} \prod_{i=1}^{k/2} \underbrace{\frac{\frac{k}{2} + i}{\ell - k + i}}_{\text{decreasing with } i}$$

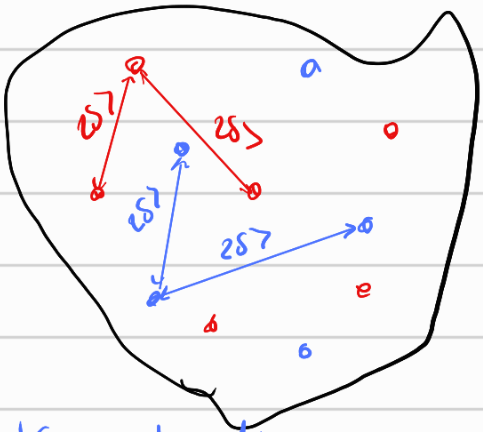
$$\leqslant \left(\frac{3}{4}\right)^{k/2} \left(\frac{\frac{k}{2} + \frac{k}{2}}{\ell - k + \frac{k}{2}}\right)^{k/2} = \left(\frac{3}{4}\right)^{k/2} \left(\frac{\ell}{k} - \frac{1}{2}\right)^{-k/2} \qquad \underbrace{k \leqslant \frac{\ell}{2}} \Rightarrow 2k \leqslant \ell \rightsquigarrow 4\ell - 2k \geqslant 3\ell$$

$$\rightsquigarrow P(A, \|\cdot\|, \tfrac{1}{2}) \geqslant \left(\frac{4}{3}\right)^{k/2} \left(\frac{\ell}{k} - \frac{1}{2}\right)^{k/2} = \left(\frac{4\ell - 2k}{3k}\right)^{k/2} \geqslant \left(\frac{3\ell}{3k}\right)^{k/2} = \left(\frac{\ell}{k}\right)^{k/2}$$

$$\rightsquigarrow \boxed{P(A, \|\cdot\|, \tfrac{1}{2}) \geqslant \left(\frac{\ell}{k}\right)^{k/2}}$$

used methods in "How well can we estimate a sparse vector", Emmanuel et al., arXiv 1104.5246

phse 4:



$\Upsilon_u = \{x_{1,u}, \dots, x_{M,u}\}$: some 2δ packing

π_u : prior on $u \in U$

$J \in \{1, \dots, M\}$: uniformly chosen. $(J,U) \rightsquigarrow X_{J,U} = X \rightsquigarrow Y = AX + Z$

$I(X;Y|U) = \mathbb{E}_{U \sim \pi_u}[I(X;Y|U=u)]$

for fixed u: (assuming that by distance we mean the squared euclidian distance)

$$\mathcal{E} = \min_{\hat{x}} \max_{x} E[\|\hat{x} - x\|^2] \geq \min_{\hat{x}} \max_{x \in \Upsilon_u} E[|\hat{x} - x|^2] \geq \min_{\hat{x}} \frac{1}{M} \sum_{j=1}^{M} E[|\hat{x} - x_{j,u}|^2]$$

$$\geq \min_{\hat{x}} \frac{1}{M} \sum_{j=1}^{M} E[|\hat{x} - x_{j,u}|^2 \times \mathbb{1}\{|\hat{x} - x_{J,u}|^2 \geq \delta^2\}] \geq \min_{\hat{x}} \frac{1}{M} \sum_{j=1}^{M} \delta \, P[|\hat{x} - x_{J,u}|^2 \geq \delta^2]$$

$$= \min_{\hat{x}} \delta \times P_{J \sim Unif[M]}[\|\hat{x} - x_{J,u}\|^2 \geq \delta^2] \geq \min_{\hat{x}} \delta^2 \times P[\hat{J} \neq J | U=u] \quad \text{for some } u \in U$$

fano inequality: $\mathcal{E} \geq \min_{\hat{x}} \delta P[\hat{J} = J | U=u] \geq \delta^2 \left(1 - \frac{I(J;Y|U=u) + \log 2}{\log M}\right)$

$$\forall u \in U : \mathcal{E} \geq \delta^2 \left(1 - \frac{I(J;Y|U=u) + \log(2)}{\log M}\right) \Rightarrow \mathcal{E} \geq \delta^2 \left(1 - \frac{\mathbb{E}_{U \sim \pi_u}[I(J;Y|U=u)] + \log 2}{\log M}\right)$$

$$\Rightarrow \mathcal{E} \geq \delta^2 \left(1 - \frac{I(J;Y|U) + \log 2}{\log M}\right)$$

$$I(J,U;Y) = I(U;Y) + I(J;Y|U) \Rightarrow I(J;Y|U) \leq I(J,U;Y)$$

$(J,U) \to X \to Y$: DP inequality: $I(J,U;Y) \leq I(X;Y)$ (I)

$$\Rightarrow I(J;Y|U) \leq I(J,U;Y) \leq I(X;Y)$$

$$\boxed{\mathcal{E} \geq \delta^2 \left(1 - \frac{I(J;Y|U) + \log 2}{\log M}\right) \overset{(I)}{\geq} \delta^2 \left(1 - \frac{I(X;Y) + \log 2}{\log M}\right)}$$

we use the $\frac{1}{2}$, norm^2 packing from phase 3 $\Rightarrow \|v_i - v_j\|_2^2 \geq \frac{1}{2} \Rightarrow \|v_i - v_j\|_2 \geq \frac{1}{\sqrt{2}}$

final phase:

$$E \geq \delta^2\left(1 - \frac{I(X;Y) + \log(2)}{\log M}\right)$$, chose points using: $X_i = c v_i$, v_i: $\frac{1}{2}$ packing $\mathcal{M}$

$M \geq \left(\frac{d}{k}\right)^{k/2} \Rightarrow \log M \geq \frac{k}{2}\log\left(\frac{d}{k}\right)$; $\|X_i - X_j\|_2 = c\,\|v_i - v_j\|_2 \geq \frac{c}{\sqrt{2}} \Rightarrow$ let $c = 2\sqrt{2}\,\delta$

$$\mathbb{E}[\|Y\|_2^2] = \mathbb{E}[\|Ax + z\|_2^2] = \mathbb{E}[x^TA^TAx + 2(Ax)^Tz + z^Tz] = \mathbb{E}[x^TA^TA\,x] + \mathbb{E}[z^Tz]$$

well known.

$$\Rightarrow \mathbb{E}[\|Y\|_2^2] = n\sigma^2 + \mathbb{E}[x^TA^TAx] \leq n\sigma^2 + \mathbb{E}\|x\|_2^2\,\sigma_{max}^2(A^TA) \leq n\sigma^2 + \mathbb{E}\,\|x\|_2^2\,\|A\|_F^2$$

$= n\sigma^2 + 8\delta^2\|v_i\|_2^2\|A\|_F^2 = n\sigma^2 + 8\delta\|A\|_F^2 \Rightarrow$ second moment at most

$t := n\sigma^2 + 8\delta^2\|A\|_F^2$

$$I(X;Y) = h(Y) - h(Z) \leq \frac{n}{2}\log\left(\frac{2\pi t e}{n}\right) - \frac{n}{2}\log(2\pi\sigma^2 e)$$

$$\Rightarrow I(X;Y) \leq \frac{n}{2}\log\left(\frac{t}{n\sigma^2}\right) = \frac{n}{2}\log\left(1 + \frac{8\delta^2\|A\|_F^2}{n\sigma^2}\right) \leq \frac{\cancel{n}}{2}\cdot\frac{8\delta^2\|A\|_F^2}{\cancel{n}\sigma^2} = \frac{4\delta^2\|A\|_F^2}{\sigma^2}$$

$$\Rightarrow E \geq \delta^2\left(1 - \frac{I(X;Y) + \log(2)}{\log(M)}\right) \geq \delta^2\left(1 - \frac{1}{\log(M)}\left[\log(2) + \frac{4\delta^2\|A\|_F^2}{\sigma^2}\right]\right)$$

$$\geq \delta\left[1 - \underbrace{\frac{2}{k\log\left(\frac{d}{k}\right)}\left(\log(2) + \frac{4\delta^2\|A\|_F^2}{\sigma^2}\right)}_{1/2}\right] \Rightarrow E \geq \frac{1}{2}\delta_{1/2}$$

let: $$\frac{1}{2} = \frac{2}{k\log\left(\frac{d}{k}\right)}\left(\log(2) + \frac{4\delta^2\|A\|_F^2}{\sigma^2}\right) \Rightarrow \delta_{1/2}^2 = \frac{\sigma^2}{4\|A\|_F^2}\left[\frac{1}{4}k\log\left(\frac{d}{k}\right) - \log(2)\right]$$

$$\boxed{\Rightarrow E \geq \frac{\sigma^2}{8\|A\|_F^2}\left[\frac{1}{4}k\underbrace{\log\left(\frac{d}{k}\right)}_{\geq \log(2)} - \log(2)\right] \approx C\cdot\frac{k\sigma^2}{\|A\|_F^2}\log\left(\frac{d}{k}\right)}$$