نازلان — پاسخ درس ITSL — محمدرضا دینی — ۴۰۰۱۰۱۲۰۴ — سوال 2: — به نام ایزد یکتا

(I) $Y \sim \mathcal{N}(0, \beta I) \in \mathbb{R}^{\ell}$

$$D_{KL}(P_X \| P_Y) = \mathbb{E}_{X\sim P_X}\left[-\log\left(\frac{P_Y}{P_X}\right)\right] = \mathbb{E}_{X\sim P_X}\left[-\log\left(\frac{1}{P_X}\right) + \log P_Y\right] = -h(X) + \mathbb{E}_{X\sim P_X}\left[\log P_Y\right]$$

$$= -h(X) + \mathbb{E}_{X\sim P_X}\left[\frac{\ell}{2}\log 2\pi\beta + \frac{1}{2\beta}\|X\|_2^2\right] = -h(X) + \frac{\ell}{2}\log 2\pi\beta + \frac{1}{2\beta}\mathbb{E}\left[\|X\|_2^2\right]$$

(II) let $\mathbb{E}\left[\|X\|_2^2\right] \le t$ since from (I) we get:

$h(Y) = -D_{KL}(P_X \| P_X) + \frac{\ell}{2}\log(2\pi\beta) + \frac{1}{2\beta}\mathbb{E}\left[\|X\|_2^2\right]$. In order to maximize $h(Y)$, one can increase $\mathbb{E}\|X\|_2^2$ to its max, & also decrease $D_{KL}(P_X \| P_X)$ since $D_{KL} \ge 0$ its min $= 0$, therefore greedy thinking we will get $\mathbb{E}\left[\|X\|^2\right] = t$, $P_X = P_Y$ or $D_{KL}(P_X \| P_X) = 0$. therefore Y must take Normal distribution (its mean (μ) plays no role. so let it be 0.) $\Rightarrow$ $Y \sim \mathcal{N}(0, \beta I)$

$$\frac{\partial h(Y)}{\partial \beta} = \frac{\partial}{\partial \beta}\left[\frac{\ell}{2}\log(2\pi\beta) + \frac{1}{2\beta}\underbrace{\mathbb{E}\left[\|X\|^2\right]}_{t}\right] = 0$$

$$\hookrightarrow \frac{2\pi\ell}{4\pi\beta} + \frac{-t}{2\beta^2} = 0 \rightarrow \frac{t}{2\beta^2} = \frac{\ell}{2\beta} \rightarrow \boxed{\beta^* = \frac{t}{\ell}}$$

thus: $h_{max}(Y) = \frac{\ell}{2}\log(2\pi\beta) + \frac{t}{2\beta} = \frac{\ell}{2}\log\left(\frac{2\pi t}{\ell}\right) + \frac{\ell}{2} = \boxed{\frac{\ell}{2}\log\left(\frac{2\pi t e}{\ell}\right)}$

(III) since $I(X;Y) = h(X) - h(Y|X)$ & the fact that $Y = AX + Z$, we can deduce that $h(Y|X) = h(Z)$, therefore $\boxed{I(X;Y) = h(Y) - h(Z)}$

Q.E.D. ■

1 نمره

Scanned by CamScanner

(I) We Can see that the output of multiplication of two Signed - :پاسخ
Permutation Matrices A, B ; are always in $\{0, \pm 1\}$. So we must just Prove that in each row & each column of C, there is only one nonzero element.

We know that each row & column of A, B is one of the $\{\pm e_1, \pm e_2, \dots, \pm e_\ell\}$, so when considering $A b_i = \begin{bmatrix} a_1 & \cdots & a_\ell \end{bmatrix}^T b_i$, there will be one and exactly one row of A (let it be a_k) that $a_k^T b_i \neq 0$. And this will prove that each column of C can not be zero and it will be in the set $\{\pm e_1, \dots, \pm e_\ell\}$.

We Can do the same thing for $a_i^T B = a_i^T \begin{bmatrix} b_1 & \cdots & b_\ell \end{bmatrix}$ and it will be suffice to Prove it.

2 نمره

So the multiplication of two signed - Permutation Matrices will be a Signed - Permutation matrix as its elements $c_{ij} \in \{0, \pm 1\}$ & each row & column has exactly one non-zero element. Q.E.D ∎

Scanned by CamScanner

(II) let x be k-sparse vector ($\|x\|_0 = k$) & B is a Signed-Permutation Matrix. :(اثبات ادعا)

Since $\forall i, \exists j : y_i = b_i^T x = \pm e_j^T x = \pm x_j$, we can deduce that $y = Bx$ has the exact elements with ~~different~~ the same absolute value & in a different order since a permutation matrix was applied.

therefore, $Y = BX$ and X have the exact number of zero elements.

therefore $Y = BX$ is a k-sparse vector since $\|x\|_0 = \|Y\|_0 \leq k$. ∎

Q.E.D

(III) Since each row of B is in the set $\{\pm e_1, \dots, \pm e_\ell\}$, then we have

$$b_i^T b_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \text{. so let } C = BB^T \rightarrow C_{ij} = b_i^T b_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

therefore we can see that $BB^T = I$, QED ∎ $B^T = B^{-1}$ $\rightarrow$ also $B^T B = I$ for the same rationale.

(IV) $\|x - x'\|_2^2 = (x - x')^T (x - x') = x^T x + x'^T x' - 2x'^T x.$

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

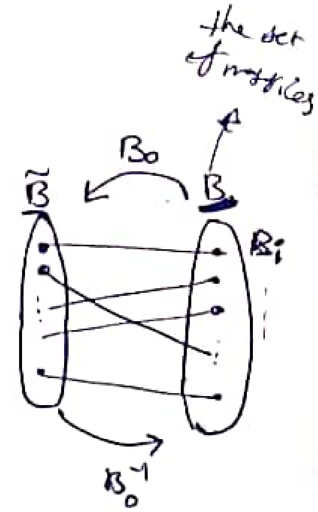
so $\|x - x'\|_2^2 = \|Bx - Bx'\|_2^2 = x^T x - x'^T x' - 2x^T x'$ Q.E.D ∎

$\|\cdot\| \geq 0 \hookrightarrow \|x - x'\|_2 = \|Bx - Bx'\|_2$

(V) Permutation Matrices $\rightarrow \ell!$ with $\{0, \pm 1\}$ elements

Signed-Permutation Matrices $\rightarrow 2^\ell \times \ell!$

(VI) So in this graph which is depicted down below we can see that each B_{0i} will correspond to exactly one element of $\tilde{B}$, since B, B_0 are full rank, they have inverse matrices, therefore $\tilde{B}$ can correspond to only one B_{0i} via B_0^{-1}. Now since the distribution of B_0's was uniform over B_0, we can deduce that $\tilde{B}$ will have uniform distribution over its elements, since we have an injective transform from B_0 to $\tilde{B}$



Q.E.D ∎

$$\frac{1}{|B|} = f_{B_0}(B) = f_{B_0}(B_0^{-1} \tilde{B}) = g_{\tilde{B}}(\tilde{B}) = \frac{1}{|B|}$$

Scanned by CamScanner

(VI) In this B is a ~~random~~ matrix. It's obvious that $\mathbb{E}[b_{ij}]=0$ for each element of B. let $y=Bx$, then

$$y_i = b_i^T x \Rightarrow \mathbb{E}[y_i] = \mathbb{E}[b_i^T x] = \mathbb{E}\Big[\sum_{j=1}^{\ell} b_{ij} x_j\Big] = \sum_{j=1}^{\ell} \underbrace{\mathbb{E}[b_{ij}]}_{0} x_j = 0$$

so $\forall i$ $\mathbb{E}\, y_i = 0$ which leads to $\mathbb{E}[y] = \mathbb{E}[Bx] = 0$.

~~For the second part we can realize that each element which does not lie on the main diagonal is zero (?)~~

4נק'

~~let $i \neq j \to c_{ij} = b_i^T x x^T b_j = \sum_{k=1}^{\ell} b_{ik} B_{kj}$~~

~~$C = Bxx^TB^T$~~

let $C = \mathbb{E}[Bxx^TB^T]$

$B_0 B_0^T = I$

Since from (V) we got that applying a matrix B to B will not change its distribution & expectation value since B_0 is a full rank matrix. so by multiplying C from both sides by B_0 we will get:

$$B_0 C B_0^T = B_0\, \mathbb{E}[Bxx^TB^T] B_0^T = \mathbb{E}\Big[\underbrace{B_0 B}_{\tilde{B}} x x^T \underbrace{B^T B_0^T}_{\tilde{B}^T}\Big] = \mathbb{E}[\tilde{B} x x^T \tilde{B}] = C$$

$B \sim \mathrm{Unif}(\frac{1}{|B|})$ $\qquad \tilde{B} \sim \mathrm{Unif}(\frac{1}{|B|})$

Since $B_0^T = B_0^{-1}$ and $B_0 C B_0^T = C \to B_0 C \underbrace{B_0^T B}_{I} = C B_0 \to C B_0 = B_0 C$

Main Proof:

(*) now will prove that for each C_{ij} where $i \neq j$; $C_{ij} = 0$.

$$c_{ij} = b_i^T b_j (x_i x_j) = x_i x_j \times b_i^T b_j = x_i \cdot x_j \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases} = x_i^2$$

$b_i \to$ row

now since there is symmetry we can tell that on the diameters each $x_i \in \{x_1, \ldots, x_\ell\}$ ~~...~~ has the same probability so the expected value of each C_{ii} will be: $\quad I = i \sim \mathrm{Unif}(\{1,2,\ldots,\ell\})$

$$\mathbb{E}[c_{ii}] = \frac{1}{\ell} \sum_{i=1}^{\ell} x_i^2 = \frac{1}{\ell} \|x\|_2^2$$

$$i \neq j: \ \mathbb{E}[c_{ij}] = 0$$

$$C = \mathbb{E}[Bxx^TB^T] = \alpha I$$

where $\alpha = \frac{\|x\|_2^2}{\ell}$

Q.E.D ■

Scanned by CamScanner

تاریخ:

<< In this phase we used the method of a paper by Emmanuel J. Candès & Mark A. Davenport: [[How well can we estimate a sparse vector]] >>

$\|X\|_0 \le K$

if $X_i \in \{0,1\} \rightarrow \|X\| \le 1$

We can see that each x with k-sparsity lies into a sphere of radius 1 in d-dimensions.. so now we will find $f_X(x)$ where x can have up to 'k', nonzero elements. so let $\bar{x} = [\underbrace{1,1,\dots,1}_{k},0,\dots 0] \rightarrow$ so the first k-elements of B will matter

با حل این مقیاس تعداد عناصر

$$f_X(X) = \mathbb{P}[X=x] = \mathbb{P}[B\bar{x}=x] = \mathbb{P}\left[B \in \left\{ \begin{array}{l} B \text{ is in first} \\ k \text{ columns of } B \end{array} \right\}\right]$$

for the first k columns

$$= \frac{k!(l-k)!\, 2^{l-k}}{2^l \times l!}$$

$\rightarrow$ $l-k$ ~~element~~ columns of B does not play a role so $(l-k)! \times 2^{l-k}$ different metrices

$\rightarrow$ here the degree of freedom is up to l. $\rightarrow$ so $l! \times 2^l$ different metrices

$$f_X(X) = \frac{k!(l-k)!}{l!} \times 2^{-k} = \frac{2^{-k}}{\binom{l}{k}} = \frac{1}{2^k \binom{l}{k}}$$

since we wanted to count the number of x's into the sphere which is $2^k \times \binom{l}{k} = |A|_{k,l}$

Now we let $\varepsilon = 1/2$.

$$A_{k,l} = \left\{ x \mid x = \frac{1}{\sqrt{k}}(x_1,\dots,x_l),\ \forall i\ x_i \in \{0,\pm 1\} \right\}$$
$$\|X_0\| = k$$

let $N_\varepsilon(P)$ denote the number of points with distance at most ε with respect to P. now if we sum all the points in A, we will ensure that this number exceeds $|A|$. thus (let us assume that the points are positioned optimaly. these points cover all B with $\varepsilon = \frac{1}{2}$)

$$\sum_{P \in B} N_\varepsilon(P) \ge |A|$$

$$|B| = P(A, \|\cdot\|_2, \varepsilon = \tfrac{1}{2}), \quad B \subseteq A$$

$$\hookrightarrow P(A, \|\cdot\|_2, \varepsilon) \times |\mathcal{N}_\varepsilon| \ge |A| \Rightarrow P(A, \|\cdot\|_2, \varepsilon) \ge \frac{|A|}{|\mathcal{N}_\varepsilon|}$$

ince $\mathcal{N}_\varepsilon = \{ \hat{x} \mid \hat{x} \in A,\ \|\hat{x}-x\|_2^2 \le \varepsilon \} \subseteq \{ \hat{x} \mid \hat{x} \in A,\ \frac{1}{k}\|x-\hat{x}\|_0 \le \varepsilon \}$ چون

what we did was to use the inequality to manipulate norm "2" to norm "0".

لدفعه برای اثبات این نامساوی
راه حل دیگری نیست را فوق را باید بزنیم.

صفحه 5

Scanned by CamScanner

where $\mathcal{U} = \{x \in \{0, +\frac{1}{\sqrt{k}}, -\frac{1}{\sqrt{k}}\}^d, \|x\|_0 = k\}$, $|\mathcal{U}| = \binom{d}{k} 2^k$

$\forall x \,\&\, x' \in \mathcal{U}$; $\frac{1}{k}\|x' - x\|_0 \leq \|x' - x\|_2^2$ & thus if $\|x' - x\|_2^2 \leq \varepsilon = \frac{1}{2}$, then $\|x' - x\|_0 \leq \frac{k}{2}$. from this we observe that for any fixed $x \in \mathcal{U}$

$$\left|\{x' \in \mathcal{U} : \|x' - x\|_2^2 \leq \varepsilon\}\right| \underset{\subseteq}{\leq} \left|\{x' \in \mathcal{U} : \|x' - x\|_0 \leq \frac{k}{2}\}\right| \leq \binom{n}{k/2} 3^{k/2}$$

each element is either $\{\pm\frac{1}{\sqrt{k}}, 0\}$

suppose we wanted to construct $\mathcal{X}$ by picking elements of $\mathcal{U}$ at random. when adding the j^{th} point point to $\mathcal{X}$ (denotably x_j), the Probability that x_j violates ~~$\forall x_j$~~ $\forall x_i, x_j \in \mathcal{X}$ $\|x_i - x_j\|_2^2 \geq \varepsilon$ is bounded by:

$$\frac{(j-1)\binom{d}{k/2} 3^{k/2}}{\binom{d}{k} 2^k} \longrightarrow \text{by union bound}$$

$$P_1 \leq \frac{|\mathcal{X}|^2}{2} \frac{\left(\binom{d}{k/2} 3^{k/2}\right)}{\binom{d}{k} 2^k}$$

so we get $\mathcal{P}(A, \|\cdot\|_2, \varepsilon) \geq \dfrac{|A|}{|N_\varepsilon|}$

$$\frac{|N_\varepsilon|}{|A|} \leq \frac{\binom{d}{k/2} 3^{k/2}}{\binom{d}{k} 2^k} = \left(\frac{3}{4}\right)^{k/2} \frac{\binom{d}{k/2}}{\binom{d}{k}} = \left(\frac{3}{4}\right)^{k/2} \frac{k!\,(d-k)!}{(\frac{k}{2})!\,(d-\frac{k}{2})!} = \left(\frac{3}{4}\right)^{k/2} \times \prod_{i=1}^{\lfloor k/2 \rfloor} \frac{\lfloor k/2 \rfloor + i}{d - k + i}$$

$$\leq \left(\frac{3}{4}\right)^{k/2} \left(\frac{k/2 + k/2}{d - k + \frac{k}{2}}\right)^{k/2} = \left(\frac{3}{4}\right)^{k/2} \left(\frac{d}{k} - \frac{1}{2}\right)^{k/2}$$

& since $k \leq \frac{d}{2} \rightarrow 2k \leq d \xrightarrow{+2d} 4d - 2k \geq 3d$

$$\mathcal{P}(A, \|\cdot\|_2, \varepsilon) \geq \left(\frac{4d - 2k}{3k}\right)^{k/2} \geq \left(\frac{3d}{9k}\right)^{k/2} \geq \left(\frac{d}{k}\right)^{k/2}$$

finally $\Rightarrow$ $\boxed{\mathcal{P}(A, \|\cdot\|_2, \varepsilon) \geq \left(\frac{d}{k}\right)^{k/2}}$ Q.E.D

ضريب

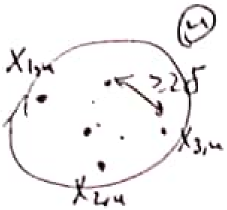
Scanned by CamScanner

$$\varepsilon = \min_{\hat{X}(Y)} \max_{X \in S_u^d} \mathbb{E}\left[\|\hat{X} - X\|^2\right]$$

for a given $U = u$: $\quad J - X - Y_{|U} \quad \hat{X}$ $\quad$:(4,b)

here we have observed $Y|U$ & estimated X. Furthermore, Instead of one ε-packing set, we have $|\mathcal{U}|$ ε-packing sets from our parametric set Θ.

let $T_u = \{X_{1,u}, \dots, X_{m,u}\}$ be the points such that they have at least 2δ distance



let us consider only this case, we know from Fano's inequality that

$$\varepsilon \geq \delta^2 \cdot \mathbb{E}_{U \sim \pi_u}\left[\mathbb{P}[J \neq \hat{J}]\right] \xrightarrow{by} \geq \delta^2 \mathbb{E}_{U \sim \pi_u}\left(1 - \frac{I(J;Y|\overset{U=u}{U}) + \log 2}{\log M}\right)$$

thus:

$$\Rightarrow \varepsilon \geq \delta^2 \mathbb{E}_{U \sim \pi_u}\left[1 - \frac{I(J;Y|U^{=u}) + \log 2}{\log M}\right] \xrightarrow{lin.} = \delta^2 \cdot 1 - \frac{\mathbb{E}_{U \sim \pi_u}\left[I(J;Y|U=u)\right] + \log 2}{\log M}$$

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

since by strong DPI we have $I(J;\underset{\text{no}}{Y}) \leq I(X;Y)$, from $(*)$ we can deduce that: $\hookrightarrow Y|U$

$$\varepsilon \geq \delta^2\left(1 - \frac{I(X;Y) + \log 2}{\log M}\right) \qquad \text{Q.E.D.} \ \blacksquare$$

$$(**) \quad \mathbb{E}_{\pi_u}\left[I(J;Y|U=u)\right] = I(J;Y|U) \leq I(X;Y)$$

$\Phi(\delta) = \delta^2$

$\hookrightarrow$ با این فرض سوال
حل شده است

صفحه 7

Scanned by CamScanner

تاریخ:

Yet we observed $E = \min_{\hat{x}(Y)} \max_{X} \mathbb{E}\left[\|X - \hat{x}(Y)\|_2^2\right]$,

E is bounded by $\delta^2\left(1 - \frac{I(X;Y) + \log 2}{\log M}\right)$. In Previous Parts we found that: (we use the things we did in Assouad method)

$$\begin{cases} I(X;Y) = h(Y) - h(Z) \\ M \geq \left(\frac{d}{k}\right)^{k/2} \end{cases} \qquad \begin{cases} \operatorname{argmax}_{Y}\{h(Y)\} = Y \sim N(0, \frac{t}{n} I) \\ \mathbb{E}\|T\|^2 \leq t \end{cases}$$

So we wish to minimize the R.H.S. by maximizing $I(X;Y)$ which is done when $h(Y)$ is maximized $\to$ $Y \sim N(0, \beta I)$.

$$\mathbb{E}\left[\|Y\|_2^2\right] = \mathbb{E}\left[\|AX + Z\|_2^2\right] = \mathbb{E}\left[X^T A^T A X + 2X^T A Z + Z^T Z\right] \qquad \mathbb{E}Z = 0$$

since $\mathbb{E}\left[Z^T Z\right] = n\sigma^2 \longrightarrow \mathbb{E}\left[\|Y\|_2^2\right] = n\sigma^2 + \mathbb{E}\left[X^T A^T A X\right]$

$$\mathbb{E}\left(\|Y\|_2^2\right) \leq n\sigma^2 + \left(\mathbb{E}\|X\|^2\right) \sigma_{max}(A^T A) \leq n\sigma^2 + \left(\mathbb{E}\|X\|_2^2\right) \|A\|_F^2$$

$$\Leftrightarrow \mathbb{E}\left[\|Y\|_2^2\right] \leq n\sigma^2 + 8\delta^2 \|V_i\|_2^2 \|A\|_F^2 = n\sigma^2 + 8\delta^2 \|A\|_F$$

the main part — توجه

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

$$\Rightarrow I(X;Y) \leq \frac{n}{2}\log\left(\frac{t}{n\sigma^2}\right) = \frac{n}{2}\log\left(1 + \frac{8\delta^2\|A\|_F^2}{n\sigma^2}\right) \xrightarrow[\text{if } x > 0]{\log(1+x) \leq x}$$

$$\Rightarrow I(X;Y) \leq \frac{n}{2}\left(\frac{8\delta^2\|A\|_F^2}{n\sigma^2}\right) = \frac{4\delta^2\|A\|_F^2}{\sigma^2} \qquad (*)$$

now plugging $(*)$ into the main assertion from 4th phase implies that

$$E \geq \delta^2\left(1 - \frac{I(X;Y) + \overset{\to \text{ dump}}{\log 2}}{\log M}\right) \geq \delta^2\left(1 - \underbrace{\frac{2}{k\log(\frac{d}{k})}\left(\frac{4\delta^2\|A\|_F^2}{\sigma^2}\right)}_{\text{let } = \frac{1}{2}}\right)$$

Scanned by CamScanner

الرقم فاضي

$$c = o(1)$$

$$\frac{1}{2} = 1 - \frac{2}{k\log(\frac{d}{k})}\left(\frac{4\delta^2\|A\|_F^2}{\sigma^2}\right)$$

$$\hookrightarrow 4\delta^2\|A\|_F^2 = 4k\log(\tfrac{d}{k}) \rightarrow \delta^2 = \frac{k\log(\frac{d}{k})\sigma^2}{\|A\|_F^2} \times c$$

$$\boxed{E \geq \delta^2 \times \frac{1}{2} = \frac{c\,k\log(\frac{d}{k})\sigma^2}{\|A\|_F^2}}$$

Q.E.D ■

صفحة 9

Scanned by CamScanner