HW2 — تمرین سری 2 آنالیز احتمالاتی ابعاد بالا : محمد عارف دینی 400101204

سوال 1.

(I) Since the set of points in an interior covering is a subset of points in an exterior covering we can deduce that $N^{ext}(K,d,\varepsilon) \leq \underset{(exterior)}{N}(K,d,\varepsilon)$. $\boxed{(*)}$

For the other direction let $A=\{a_1, \dots, a_m\}$ be an $\frac{\varepsilon}{2}$ covering of the set K. then $K \subseteq \bigcup_{i=1}^{m} B_{\varepsilon/2}(a_i)$. Now if we chose $\{b_j\}_{j=1}^{m}$ from each $B_{\varepsilon/2}(a_j) \cap K$, then:

$b_j \in B_{\varepsilon/2}(a_j) \longrightarrow \bigcup_{j=1}^{n} B_{\varepsilon}(b_j) \supseteq K$. thus we found a ε-covering with m points for the set K, which suggests that perhaps the minimal number of points needed for ε-covering (interior) must be less or equal to $m = N^{ext}(K,d,\varepsilon)$. $\boxed{(**)}$

So now combining $(*)$ & $(**)$ implies that: $\boxed{N^{ext}(K,d,\varepsilon) \leq N(K,d,\varepsilon) \leq N^{ext}(K,d,\varepsilon)}$

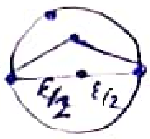
Q.E.D. ∎

(II) just let $L=\{-\varepsilon, +\varepsilon\}$, $K=[-\varepsilon, \varepsilon]$ then if $d(x,y) = |x-y|$:

$N(L,d,\varepsilon) = 2 < N(K,d,\varepsilon)$ ✗.

this time let $\{a_1, -, a_m\}$ be an $\frac{\varepsilon}{2}$-covering for the set K. then we'd say:

$K \subseteq \bigcup_{i=1}^{n} B(a_i, \frac{\varepsilon}{2}) \longrightarrow \forall y,z \in B_{\frac{\varepsilon}{2}}(a_i) \longrightarrow d(y,z) \leq \varepsilon$.



Now let $A' = \{a_1, \dots, a_{m-t}\}$ (where $t>0$) be the points such that $B_{\varepsilon/2}(a_i) \cap L \neq \phi$. the same thing we did above. we choose each b_j from $B_{\varepsilon/2}(a_j) \cap L$.

$\boxed{\forall p \in L: \exists b_j \to d(p,b_j) \leq \varepsilon} \longrightarrow \{b_1, \dots, b_{m-t}\}$ are ε-packing for the set L

$$\forall p \in L \underset{L \subseteq K}{\longrightarrow} p \in K \to \exists j: \left.\begin{array}{l} d(p,a_j) \leq \frac{\varepsilon}{2} \\ b_j \in B_{\varepsilon/2}(a_j) \cap L \end{array}\right\} d(p,b_j) \leq \varepsilon$$

Since $N(L,d,\varepsilon) \leq m-t \leq m = N(K,d,\frac{\varepsilon}{2})$ we are done! Q.E.D.

if $\boxed{L \subseteq K: N(L,d,\varepsilon) \leq N(K,d,\varepsilon/2)}$ Q.E.D. ∎

(III) Let $\{a_1, \dots, a_n\}$ where $n = P(K, d_H, m)$ be an m-packing for the set $K=\{0,1\}^n$. We can see that $\{a_1, \dots, a_n\}$ is a m-covering as well. otherwise there would be point a_{n+1} that has a distance less than m with all $\{a_i\}_{i=1}^{n}$ which can't happen due to maximality of packing $\{a_1 - a_n\}$.

thus: $N(K, d_H, m) \leq P(K, d_H, m)$ ✓ $\boxed{(**)}$

100

Scanned by CamScanner

Now we consider an m-packing $\{a_1, \dots, a_m\}$ of binary strings (of $\{0,1\}$) of length n. Now since each a_i, a_j have at least m different binary digits; if we take each a_i and alter $t \le \lfloor \frac{m}{2} \rfloor$ digits of a_i, we will still get distinct elements $\mathcal{E} = \{x : d_H(x, a_i) \le \lfloor \frac{m}{2} \rfloor, \{a_i\}_{i=1}^{m} \text{ is m-packing}\}$. Since the total number of strings of length n is 2^n we have:

$$|\mathcal{E}| = P(K, d_H, m) \cdot \left(\sum_{k=0}^{\lfloor m/2 \rfloor} \binom{n}{k}\right) \le 2^n = |K| \checkmark \quad \boxed{(**)}$$

For the other side of the inequality, let $\{b_1, \dots, b_n\}$ be an m-covering of the set $K = \{0,1\}^n$. For each b_j we change at most m binary digits, we claim that we will cover all elements of K, since:

$\forall x \in K, \exists b_j \in \{b_1, \dots, b_n\} \to d(b_j, x) \le m \to$ so for each b_j if we change m binary digits we can cover $\{0,1\}^n$.

therefore: $|K| = 2^n \le N(K, d_H, m) \cdot \left(\sum_{k=0}^{m} \binom{n}{k}\right)$ ✓ (***)

Mixing (*), (**), (***) →

$$\frac{2^n}{\sum_{k=1}^{m}\binom{m}{k}} \le N(K, d_H, m) \le P(K, d_H, m) \le \frac{2^n}{\sum_{k=0}^{\lfloor m/2 \rfloor}\binom{m}{k}}$$

Q.E.D.

(IV) it suffices to prove $\frac{1}{d+1} e^{-d \cdot D_{KL}(\frac{\delta}{2} \| \frac{1}{2})} \le \frac{1}{P(K, d_H, \delta)}$

Let $\{a_1, \dots, a_m\}$ be a δ-packing of the set $\{0,1\}^n$ with normalized Hamming distance. Now if we add a binary noise of length d. then the probability of them be equal to one of a_i's is $\frac{1}{P(\{0,1\}^d, d_H, \delta)}$. Now let the sum of the binary noise to be Z_d, Then

$$\mathbb{P}\left[Z_d \le \frac{\delta d}{2}\right] \le \frac{1}{P(\{0,1\}^d, d_H, \delta)}$$

Furthermore, we know that $\mathbb{P}\left[Z_d \le \frac{\delta d}{2}\right] \ge \frac{1}{d+1} e^{-d \cdot D(\frac{\delta}{2} \| \frac{1}{2})}$

which results in:

Q.E.D. $\boxed{\log(P(K, d_H, \delta)) \le d \cdot D_{KL}\left(\frac{\delta}{2} \| \frac{1}{2}\right) + \log(d+1)}$

$A_1,\dots,A_n$ are independent events. $\quad a\wedge b = \min(a,b)$ $\quad$ پاسخ:

(I) From Union Bound we have: $(\forall \mathcal{E}\in\Omega: \mathbb{P}[\mathcal{E}]\le 1)$ $\quad$ The Proof of the R.H.S of inequality

$$\mathbb{P}\left[\bigcup_{k=1}^{n} A_k\right] \le \mathbb{P}[A_n]+\mathbb{P}\left[\bigcup_{k=1}^{n-1} A_k\right] \le \dots \le \sum_{k=1}^{n}\mathbb{P}[A_k].\ \text{Thus: } \mathbb{P}\left[\bigcup_{k=1}^{n} A_k\right]\le 1\wedge\left(\sum_{k=1}^{n}\mathbb{P}(A_k)\right)$$

Q.E.D. ■

from the hint lemma: $\forall x: (1-e^{-1})(1\wedge x)\le 1-e^{-x}$. Now let $x=\sum_{k=1}^{n}\mathbb{P}[A_k]$:

$$(1-e^{-1})\left(1\wedge\left[\sum_{k=1}^{n}\mathbb{P}(A_k)\right]\right)\le 1-\exp\left(-\sum_{k=1}^{n}\mathbb{P}(A_k)\right).$$

furthermore, from the other lemma: $\forall x_1\dots x_n: \prod_{k=1}^{n}(1-x_k)\le \exp\left(-\sum_{i=1}^{n}x_k\right)$. Now let $\forall k: x_k=\mathbb{P}[A_k]$

$$1-\exp\left(-\sum_{i=1}^{n}\mathbb{P}(A_k)\right)\le 1-\prod_{k=1}^{n}\left(1-\mathbb{P}(A_k)\right)=1-\mathbb{P}\left[\bigcap_{k=1}^{n}\overline{A_k}\right]=\mathbb{P}\left[\bigcup_{k=1}^{n}A_k\right].\ \checkmark$$

which follows that: $\boxed{(1-e^{-1})\left\{1\wedge\left[\sum_{k=1}^{n}\mathbb{P}(A_k)\right]\right\}\le \mathbb{P}\left[\bigcup_{k=1}^{n}A_k\right]\le\left\{1\wedge\left[\sum_{k=1}^{n}\mathbb{P}(A_k)\right]\right\}}$ Q.E.D. ■

(II) η^*: Convex & increasing, $\{X_t, t\in T\}$ are independent; $\forall t\in T, x\ge 0: \mathbb{P}[X_t\ge x]\le e^{-\eta^*(x)}$.

let A_t be the event: $X_t\ge \eta^{*-1}(\log|T|+u)=S$.

$$\mathbb{P}\left[\sup_{t\in T}X_t\ge S\right]=1-\mathbb{P}\left[\sup_{t\in T}X_t\le S\right]=1-\mathbb{P}\left[\forall t\in T: X_t\le S\right]=1-\mathbb{P}\left[\bigcap_{t\in T}\overline{A_t}\right]$$

$$\rightarrow \mathbb{P}\left[\sup_{t\in T}X_t\ge S\right]=\mathbb{P}\left[\bigcup_{t\in T}A_t\right]\underset{\text{from (I)}}{\ge}(1-e^{-1})\left[\sum_{t\in T}\mathbb{P}[A_t]\right]$$

$\forall t\in T$:

$$\mathbb{P}(A_t)=\mathbb{P}[X_t\ge S]\ge \exp\left[-\eta^*\left(\eta^{*-1}(\log|T|+u)\right)\right]=\exp(-\log|T|-u)$$

which leads to: $\mathbb{P}\left[\sup_{t\in T}X_t\ge S\right]\ge(1-e^{-1})\left(\sum_{t\in T}\frac{e^{-u}}{|T|}\right)=(1-e^{-1})e^{-u}$. ■ Q.E.D.

(III) It suffices to let $u=\frac{\eta^*(2x)}{2}$, then since η^{*-1} is concave we'd have $f\left(\frac{x}{2}+\frac{y}{2}\right)\ge\frac12 f(x)+\frac12 f(y)\ \forall x,y\in\mathrm{dom}f$; Using all these we get:

$$S=\eta^{*-1}(+\log|T|+u)=\eta^{*-1}\left(+\frac{2\log|T|}{2}+\frac{\eta^*(2x)}{2}\right)\ge\frac12\eta^{*-1}(\log|T|)+\underbrace{\frac12\eta^{*-1}(\eta^*(2x))}_{x}=S'$$

$$\overset{\text{(II)}}{\rightarrow}\mathbb{P}\left[\sup_{t\in T}X_t\ge S\right]\underset{S\ge S'}{\ge}\mathbb{P}\left[\sup_{t\in T}X_t\ge S'\right]\ge(1-e^{-1})\exp\left(-\frac{\eta^*(2x)}{2}\right)$$ Q.E.D. ■

(IV) ψ^* increasing, convex. $e^{-\eta^*(x)} \le \mathbb{P}[X_t \ge x] \le e^{-\psi^*(x)} \quad \forall t \in T, x > 0$

(i) Since $X_t \ge 0 \ \forall t \in T$: then $\mathbb{E}\left[\sup_{t\in T} \min(0, X_t)\right] + \mathbb{E}\left[\sup_{t\in T} \max(0,X_t)\right] = \mathbb{E}\left[\sup_{t\in T} X_t\right]$

$$\mathbb{E}\left[\sup_{t\in T} \max(0, X_t)\right] + \mathbb{E}\left[\sup_{t\in T} \min(0,X_t)\right] \ge \sup_{t\in T}\left\{\mathbb{E}\left[\min(0,X_t)\right] + \mathbb{E}\left[\max(0,X_t)\right]\right\}$$

$$\ge \sup_{t\in T}\left\{\mathbb{E}\left[\min(0,X_t)\right] + \int_0^\infty \mathbb{P}\left[\sup_t \max(X_t,0) \ge u\right] du\right\}$$

$$\ge \sup_{t\in T}\left\{\mathbb{E}\left[\min(0,X_t)\right] + \int_0^{\eta^{*-1}(\log|T|)} \mathbb{P}\left[\sup_{t\in T} X_t \ge \eta^{*-1}(\log(|T|))\right] du\right\}$$

$$\ge \sup_{t\in T}\left\{\mathbb{E}\left[\min(0,X_t)\right] + \psi^{*-1}(\log|T|)\,(1-e^{-1})\right\}$$

$$\ge \underbrace{(1-e^{-1})}_{C_1>0}\left[\sup_{t\in T}\left\{\min(0,X_t)\right\} + \eta^{*-1}(\log|T|)\right]$$

Q.E.D. ■ ✓

(ii) $$\mathbb{E}\left(\sup_{t\in T} X_t\right) \le \mathbb{E}\left[\sup_{t\in T}\{\max(0,X_t)\}\right] = \int_0^\infty \mathbb{P}\left[\sup_{t\in T} X_t \ge u\right] du$$

$$= \int_0^{\psi^{*-1}(\log|T|)} \underbrace{\mathbb{P}\left[\sup_{t\in T} X_t \ge u\right]}_{\le 1} du + \int_0^\infty \mathbb{P}\left[\sup_{t\in T}\{X_t\} \ge \psi^{*-1}(\log|T|) + u\right] du$$

Union Bound

$$\le \psi^{*-1}(\log|T|) + |T|\int_0^\infty e^{-\psi^*(\psi^{*-1}(\log|T|)+u)} du$$

$|T| = e^{\psi^*(\psi^{*-1}(\log|T|))}$

$$\le \psi^{*-1}(\log|T|) + \int_0^\infty e^{-\psi^*(\psi^{*-1}(\log|T|)+u) - \psi^*(\psi^{*-1}(\log|T|))} du$$

ψ^* is convex

$$\le \psi^{*-1}(\log|T|) + \int_0^\infty e^{-\psi^*(u) - \psi^{*\prime}(\psi^{*-1}(\log|T|))(u - \psi^{*-1}(\log|T|))}$$

$\psi^*(a+x) \ge \psi^*(a) + (a-x)\psi^{*\prime}(x)$

$$\le \psi^{*-1}(\log|T|) + C = O(\psi^{*-1}(\log|T|))$$

$A \le \psi(x) \quad \forall x \ge 0$

(V) $X \sim N(0,1) \longrightarrow \mathbb{P}(X \geq x) \geq \frac{e^{-\frac{x^2}{2}}}{2\sqrt{2}}$ $\quad (x+u)^2 \leq 2u^2 + 2x^2$

$$\mathbb{P}(X \geq x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{(u+x)^2}{2}} du \leq \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{2u^2+2x^2}{2}} du$$

$$\rightarrow \mathbb{P}(X \geq x) \leq \left(\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{2u^2}{2}} du\right) e^{-x^2} = \frac{e^{-x^2}}{2\sqrt{2}}$$ Q.E.D. ■

(VI) $X_1, \dots, X_n \overset{iid}{\sim} N(0,1)$

Since $\begin{cases} \mathbb{P}(X \geq x) \leq e^{-\frac{x^2}{2}} \\ \mathbb{P}(X \geq x) \geq e^{-\frac{x^2}{2} - \log(2\sqrt{2})} \end{cases} \longrightarrow$ let $\begin{cases} \psi^*(x) = \frac{x^2}{2} \rightarrow \psi^{*-1}(x) = \sqrt{2x} \\ \eta^*(x) = \frac{x^2}{2} - \underbrace{\log(2\sqrt{2})}_{c} \rightarrow \eta^{*-1}(x) = \sqrt{2x - c} \end{cases}$

$$\mathbb{E}\left[\max_i [X_i]\right] \overset{\lambda \geq 0}{\leq} \frac{1}{\lambda} \mathbb{E}\left[\log\left(\max_i \{e^{\lambda X_i}\}\right)\right] \leq \frac{1}{\lambda} \mathbb{E}\left[\log\left(\sum_{i=1}^n e^{X_i \lambda}\right)\right]$$

$$\leq \frac{1}{\lambda} \log\left(\mathbb{E} \sum e^{\lambda X_i}\right) = \frac{1}{\lambda} \log(n \, \mathbb{E} e^{\lambda X_i}) \leq \frac{\log(n)}{\lambda} + \frac{\lambda}{2}$$

Now optimizing on λ implies that: $\lambda^* = \sqrt{2 \log n}$ & thus:

$$\mathbb{E}\left[\max_i \{X_i\}\right] \leq \sqrt{2 \log n}$$ Q.E.D. ■

from (III) we got

$$\mathbb{E}\left[\max_i X_i\right] = (1 - e^{-1}) \left[\underbrace{\sup_{t \in T} \mathbb{E}[\min(0, X_t)]}_{} + \eta^{*-1}(\log n)\right]$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{x^2}{2x}} dx = \frac{-1}{\sqrt{2\pi}}$$

$$= \frac{(1 - e^{-1})}{2} \boxed{\left[\frac{-1}{\sqrt{2\pi}} + \sqrt{2 \log(n \cdot 2^{-\frac{3}{4}})}\right]}$$ Q.E.D. ■

thus for $n \to \infty$ we get:

$$\exists c, C: \quad c\sqrt{\log n} \leq \mathbb{E}\left[\max_i \{X_i\}\right] \leq C\sqrt{\log n}$$

$$\longrightarrow \mathbb{E}\left[\max_i \{X_i\}\right] = \Theta(\sqrt{\log n})$$

(i) $\|X_{\sigma(i)} - X_{\sigma(i+1)}\| \geq \min_{l \neq \sigma(i)} \{\|X_{\sigma(i)} - X_l\|\} \rightarrow \sum_{i=1}^{n} \|X_{\sigma(i)} - X_{\sigma(i+1)}\| \geq \sum_{i=1}^{n} \min_{l \neq \sigma(i)} \{\|X_{\sigma(i)} - X_l\|\}$

$\mathcal{L}_n = \min_{\sigma} \left\{ \sum_{i=1}^{n} \|X_{\sigma(i)} - X_{\sigma(i+1)}\| \right\} \geq \sum_{i=1}^{n} \min_{l \neq k} \{\|X_k - X_l\|\}$

$\hookrightarrow \mathbb{E}[\mathcal{L}_n] \geq \sum_{k=1}^{n} \mathbb{E}\left[\min_{k \neq l} \{\|X_k - X_l\|\}\right]$ in order to get this we need to find this

X_i's are independent

on the other hand: $\mathbb{P}\left[\min_{l \neq k} \|X_k - X_l\| \geq t\right] = \mathbb{P}\left(\forall l \neq k : \|X_k - X_l\| \geq t\right] = \mathbb{P}\left[\|X_k - X_l\| \geq t\right)^{n-1}$

which results in: $\mathbb{P}\left(\min_{l \neq k} \|X_k - X_l\| \geq t\right] \geqslant (1 - \pi t^2)^{n-1} \geq (1 - \pi t^2)^n \quad \left(\forall t \in [0, \frac{1}{\sqrt{\pi}}]\right)$

. Since it only applies to $t \in [0, \frac{1}{\sqrt{\pi}}]$, for $t \notin [0, \frac{1}{\sqrt{\pi}}]$ we switch to zero:

$\mathbb{P}\left(\min_{k \neq l} \|X_k - X_l\| \geq t\right] = (1 - \pi t^2)^n \cdot \mathbb{1}\left[t \in [0, \frac{1}{\sqrt{\pi}}]\right)$

$\rightarrow \mathbb{E}\left[\min_{k \neq l} \|X_k - X_l\|\right] = \int_0^{\infty} \mathbb{P}\left(\min_{k \neq l} \|X_k - X_l\| \geq t\right) dt \geq t^* \, \mathbb{P}\left[\min_{l \neq k} \|X_k - X_l\| \geq t^*\right]$

$\geq t^* (1 - \pi t^2)^n \cdot \mathbb{1}\{t \in [0, \frac{1}{\sqrt{\pi}}]\}$ now let $t^* = \frac{1}{\sqrt{n\pi}}$

$t^* \geq 0$

we get: $\mathbb{E}\left(\min_{k \neq l} \|X_k - X_l\| \geq t\right] \geq \frac{1}{\sqrt{n\pi}} \left(1 - \left(\frac{\pi}{n\pi}\right)\right)^n = \frac{1}{\sqrt{n\pi}} \left(1 - \frac{1}{n}\right)^n \underset{(n \geq 2)}{\geq} \frac{1}{4\sqrt{n\pi}}$

finally

$\hookrightarrow \mathbb{E}[\mathcal{L}_n] \geq \sum_{k=1}^{n} \mathbb{E}\left(\min_{k \neq l} \{\|X_k - X_l\|\right] \geq \sum_{k=1}^{n} \frac{1}{4\sqrt{n\pi}} = \frac{1}{4}\sqrt{\frac{n}{\pi}} = c\sqrt{n} \quad , c = (\sqrt{4\pi})^{-1}$ ■ Q.E.D.

(ii) $\sigma^* = \arg\min_{\sigma} \left\{ \sum_{k=1}^{n-1} \|X_{\sigma(k)} - X_{\sigma(k+1)}\| \right\}$, $i^* = \arg\min_i \{\|X_{\sigma^*(i)} - X_n\|\}$

$\mathcal{L}_n \leq \mathcal{L}_{n-1} + \|X_{\sigma^*(n-1)} - X_n\| = \sum_{k=1}^{n} \|X_{\sigma^*(i)} - X_{\sigma^*(i+1)}\| + \|X_{\sigma^*(n-1)} - X_n\|$

However, since $\|X_n - X_{\sigma^*(i^*+1)}\| \leq \|X_n - X_{\sigma^*(i^*)}\| + \|X_{\sigma^*(i^*)} - X_{\sigma^*(i^*+1)}\|$

$\rightarrow \mathcal{L}_n \leq \sum_{k=1}^{n} \|X_{\sigma^*(i)} - X_{\sigma^*(i+1)}\| + 2\|X_{\sigma^*(i^*)} - X_n\| = \mathcal{L}_{n-1} + 2\min_{k<n} \|X_k - X_n\|$ (Triangle Inequality)

$\mathcal{L}_n \leq \mathcal{L}_{n-1} + 2\min_{k<n} \|X_k - X_n\|$

$\mathcal{L}_{n-1} \leq \mathcal{L}_{n-2} + 2\min_{j<n-1} \|X_j - X_{n-1}\|$

$\vdots \quad \vdots$

$\mathcal{L}_n \leq 2\sum_{k=1}^{n} \min_{l<k} \|X_k - X_l\|$

$\mathbb{E}(\mathcal{L}_n) \leq 2\sum_{k=1}^{n} \mathbb{E}\left[\min_{k>l} \{\|X_k - X_l\|\}\right]$

since $\mathbb{P}\left(\min_{l<k} \|X_k - X_l\| \geq t\right] \leq \mathbb{P}\left(\forall l<k : \|X_k - X_l\| \geq t\right) = \mathbb{P}\left[\|X_k - X_l\| \geq t\right]_{l \neq k}^{k-1}$

the worst case is this

this works for $0 < t < 1$ $\quad \leq \left(1 - \frac{\pi t^2}{4}\right)^{k-1}$

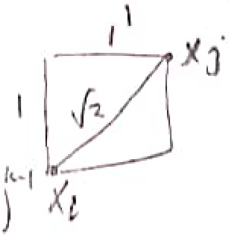
40

ادامه سوال 3:

$$\mathbb{P}\Big[\min_{l<k} \|X_k - X_l\| \ge t\Big] \le \Big(1-\frac{\pi}{4}t^2\Big)^{k-1} \mathbb{1}\{t\in(0,1)\} + \Big(1-\frac{\pi}{4}\Big)\mathbb{1}(t>1)\}$$

$$\mathbb{E}\Big[\min_{l<k} \|X_k - X_l\|\Big] = \int_0^\infty \mathbb{P}\Big[\min_{l,k} \|X_k - X_l\| \ge t\Big]dt \le \int_0^1 \Big(1-\frac{\pi}{4}t^2\Big)^{k-1}dt + \int_1^{\sqrt2}\Big(1-\frac{\pi}{4}\Big)^{k-1}dt$$

the maximum distance in a unit square is $\sqrt2$.



$$\mathbb{E}\Big[\min_{l<k} \|X_k - X_l\|\Big] \le \int_0^{2/\sqrt{n\pi}} \Big(1-\frac{\pi t^2}{4}\Big)^{k-1}dt \int_{\frac{2}{\sqrt{n\pi}}}^1 \Big(1-\frac{\pi t^2}{4}\Big)^{k-1}dt + (\sqrt2-1)\Big(1-\frac{\pi}{4}\Big)^{k-1}$$

$$\le \frac{2}{\sqrt{n\pi}} + \Big(1-\frac{1}{k}\Big)^{k-1} + (\sqrt2-1)\Big(1-\frac{\pi}{4}\Big)^{k-1} \le \frac{C}{\sqrt k} \Longrightarrow \mathbb{E}L_{(n)} \le 2\sum_{k=1}^{n}\frac{C}{\sqrt k} \le 6C\sqrt n$$

so $\exists c, C : c\sqrt n \le \mathbb{E}(L_n) \le C\sqrt n \longrightarrow \boxed{\mathbb{E}(L_n) = \Theta(\sqrt n)}$ Q.E.D. ∎

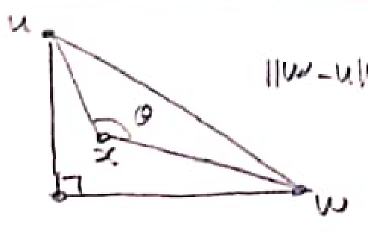
(II) let $L_n = f(X_1 - X_n) = \sum_{k=1}^{n} \|X_{\sigma^*(k)} - X_{\sigma^*(k+1)}\|$

$$f(X_1, \dots, X_j', \dots, X_n) \le \sum_{\substack{k=1\\ k\ne j, j-1}}^{n} \|X_{\sigma^*(k)} - X_{\sigma^*(k+1)}\| + \|X_j' - X_{\sigma^*(j+1)}\| + \|X_{\sigma^*(j-1)} - X_j'\|$$

$$\to f(X_{1:n}) - f(X_{1:j-1}, X_j', X_{j+1:n}) \le \|X_{\sigma^*(k+1)} - X_j'\| - \|X_{\sigma^*(k-1)} - X_{\sigma^*(k)}\| + \|X_j' - X_{\sigma^*(k+1)}\| - \|X_{\sigma^*(k)} - X_{\sigma^*(k+1)}\|$$

$$\to f(X_{1:n}) - f(X'_{1:n}) \le 2\|X_j' - X_j\|$$ $\longrightarrow$ Now According to Mcdiarmid, L_n is $\sigma^2 = \frac{n}{4}(2\sqrt2)^2 = 2n$ subgaussian.

(III)



$$\|w-u\|^2 = \|w-x\|^2 + \|u-x\|^2 + 2\underbrace{\|w-x\|\,\|u-x\|}_{\ge 0}\underbrace{\cos\theta}_{\le 0} \qquad \theta\in[\pi/2, \pi]$$

$\cos\theta \in [-1, 0]$

$$\to \boxed{\|w-u\|^2 \le \|w-x\|^2 + \|u-x\|^2}$$

We use induction on n:

for each right triangle S, $x_1, \dots, x_{n-1} \in S$, $\exists \sigma : \{1, -, n-1\} \to \{1, -, n-1\}$

such that $\|v - x_{\sigma(1)}\|^2 + \sum_{i=1}^{n} \|x_{\sigma(i)} - x_{\sigma(i+1)}\|^2 + \|x_{\sigma(m)} - w\|^2 \leq \|u - w\|^2$

Case I) there's atleast one point in both ①, ②, then

$$\|w - x_{\sigma(1)}\|^2 + \sum_{i=1}^{m-1} \|x_{\sigma(i)} - x_{\sigma(i+1)}\|^2 + \|x_{\sigma(m)}\|^2 \leq a^2$$

$$\|v - x_{\tau(n)}\|^2 - \sum_{j=m+1}^{n} \|x_{\tau(j)} - x_{\tau(j+1)}\|^2 + \|x_{\tau(m+1)}\|^2 \leq b^2$$

$$\|v - x_{\sigma(1)}\|^2 + \|w - x_{\tau(n)}\|^2 + \|x_{\sigma(m)}\|^2 + \|x_{\tau(m+1)}\|^2 + \sum_{j=m+1}^{n} \|x_{\tau(j)} - x_{\tau(j+1)}\|^2 + \sum_{i=1}^{m} \|x_{\sigma(i)} - x_{\sigma(i+1)}\|^2$$

$$\leq a^2 + b^2 = \|w - u\|^2$$

Now since $x_{\tau(m+1)}, x_{\sigma(m)} \in S \to \|x_{\sigma(m)}\|^2 + \|x_{\tau(m+1)}\|^2 \geq \|x_{\tau(m+1)} - x_{\sigma(m)}\|^2$

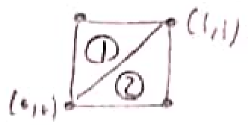
which results in $\|v - x_{\rho(1)}\|^2 + \sum_{i=1}^{n-1} \|x_{\rho(i)} - x_{\rho(i+1)}\|^2 + \|w - x_{\rho(n)}\|^2 \leq \|u - w\|^2$

Now the induction step is complete & its true because the base of induction is true as well.

Q.E.D. ■

the other Case II) is ok, since $a = 0$.

(V) let $\{x_i\}_{i=1}^{m} \in ①$, $\{x_i\}_{i=m+1}^{n} \in ②$



$\exists \sigma : \{1, -, n\} \to \{1, -, n\} : \|x_{\sigma(1)}\|^2 + \sum_{i=1}^{m-1} \|x_{\sigma(i)} - x_{\sigma(i+1)}\|^2 + \|x_{\sigma(m)} - (1,1)^2\|^2 \leq \sqrt{2}^2 = 2$

Now from cosine law.

$$\|x_{\sigma(m+1)} - x_{\sigma(m)}\|^2 \leq \|x_{\sigma(m)} - [1,1]\|^2 + \|x_{\sigma(1)} - (1,1)^2\|$$

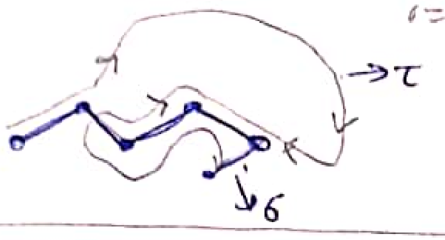
therefore:

$$\|x_{\sigma(1)}\|^2 + \sum_{i=1}^{n-1} \|x_{\sigma(i)} - x_{\sigma(i+1)}\|^2 + \|x_{\sigma(n)}\|^2 \leq 4$$

Q.E.D ■

(VI) Since $X \cap Y \neq \phi$ Some Points of Permutations σ, τ are common. We start from one of the points, we take the path of σ such that it's intersectible with the path τ. when the path got out, we'll get out as well and continue until we reach another point of τ. then we'll get back to the starting point, and go through the same path we did before. This path has a length of $2\sum_{i=1}^{n} \mathbb{1}\{X_i \notin Y\} d_i(x,\sigma)$. Thus:

$$l_{2n}(X \cup Y, \rho) \leq l_n(Y, \tau) + 2\sum_{i=1}^{n} \mathbb{1}\{x_i \notin Y\} d_i(x,\sigma) \quad \underline{Q.E.D.}$$



(VII) we change the indices such that the joint points of X, Y, have the same indices. Thus $x_i \notin Y$ will be equivalent to $x_i \neq y_i$.

now let $\tau = \arg\min_{\sigma} l_n(y,\sigma)$, $l_n(y,\tau) = \min_{\sigma} l_n(y,\sigma)$

$$\hookrightarrow \min_{\sigma} l_n(y,\sigma) \leq l_{2n}(X \cup Y, \rho)$$

$$\min_{\sigma} l_n(Y,\sigma) \leq l_{2n}(X \cup Y, \rho) \leq \min_{\sigma} l_n(Y,\sigma) + \sum_{i=1}^{n} 2d_i(Y,\sigma_x) \mathbb{1}\{x_i \neq y_i\}$$

let $f(X_1 - X_n) = \arg\min_{\sigma} l_n(X,\sigma)$

$$\rightarrow \min_{\sigma} l_n(x,\sigma) - \min_{\sigma} l(y,\sigma) \leq \sum_{i=1}^{n} 2d_i(x,\sigma_x) \mathbb{1}\{x_i \neq y_i\}$$

$$f(x) - f(y) \leq \sum_{i=1}^{n} \underbrace{2d_i(x,\sigma_x)}_{c_i(x)} \mathbb{1}(x_i \neq y_i) \rightarrow \sum_i c_i(x) = 4\sum_{i=1}^{n} d_i(x,\sigma_x) \leq 4 \times 4 = 16$$

Now According to Talagrand Inequality, $f(X)$ is subgaussian with $\sigma^2 = 16$.

$\underline{Q.E.D.}$

Scanned by CamScanner

$A \in \mathbb{R}^{m\times n}$, $A_i = [A_{i1}\ A_{i2} \cdots A_{in}] \overset{\text{indep}}{\sim} \text{SubG}$:4 OK?

the covering number of unit Euclidean ball satisfies:

$$\left(\frac{1}{\varepsilon}\right)^n \le N(B_2^n, \varepsilon) \le \left(\frac{2}{\varepsilon}+1\right)^n.$$

Thus, we can find a $\frac{1}{4}$-net covering of the unit sphere with less than 9^n points.

$\|A\| = \sup\{\|Ax\|\} \ge \sup_{x\in N}\{\|Ax\|\}$, $\|A\| \le \frac{1}{1-\varepsilon} \sup_{x\in N}\{\|Ax\|\}$

$\to$ It follows that:

$$\left\|\frac{1}{m}A^TA - I_n\right\| \le 2\max_{x\in N}\left|\left\langle\left(\frac{1}{m}A^TA - I_m\right)x, x\right\rangle\right| = 2\max_{x\in N}\left|\frac{1}{m}\|Ax\|_2^2 - 1\right|$$

Now it suffices to show that:

$$\max_{x\in N}\left\{\left|\frac{1}{m}\|Ax\|_2^2 - 1\right|\right\} \le \frac{\varepsilon}{2} \longrightarrow \varepsilon = K^2\max\{\delta, \delta^2\}$$

Now fix $x \in S^{n-1}$ & express $\|Ax\|_2^2$ as a sum of independent random variables

$$\|Ax\|_2^2 = \sum_{i=1}^n \langle A_i, x\rangle^2 = \sum_{i=1}^n Y_i^2, \quad Y_i = \langle A_i, x\rangle, \quad \mathbb{E}Y_i^2 = 1$$

Since A_i denotes the i-th row of the matrix which are independent, isotropic & subgaussian random vector with $\|A_i\|_{\psi_2} \le K$, therefore $Y_i^2 - 1$ are independent mean zero and sub-exponential. Thus using Bernstein inequality, we have: $\mathbb{P}\left[\left|\frac{1}{m}\|Ax\|_2^2 - 1\right| \ge \frac{\varepsilon}{2}\right] = \mathbb{P}\left[\left|\frac{1}{m}\sum_{i=1}^m Y_i^2 - 1\right| \ge \frac{\varepsilon}{2}\right]$

$$\le 2\exp\left[-c_1 m \min\left(\frac{\varepsilon^2}{K^4}, \frac{\varepsilon}{K^2}\right)\right] = 2\exp(-c_1 \cdot m \cdot \delta^2)$$

$$\le 2\exp\left(-c_1 \cdot c^2 (n+t)^2\right)$$

Now using Union Bound, we have:

$$\mathbb{P}\left[\max_{x\in N}\left\{\left|\frac{1}{m}A^TA - I_m\right| \ge \frac{\varepsilon}{2}\right\}\right] \le 2^n \cdot e^{-c_1 c(n+t^2)} \le 2e^{-t^2}$$

$$\longrightarrow \mathbb{P}\left[\left\|\frac{1}{m}A^TA - I_m\right\| \le \varepsilon\right] \ge 1 - 2e^{-t^2}$$

Q.E.D. ■

$\underline{X} \sim N(\vec{0}, C)$, $\underline{Y} = (Y_1, \dots, Y_n)^T \sim N(0, I)$, $\underline{X} = C^{1/2}\underline{Y}$: $\underline{5 \cup}$

let $\Sigma^{1/2} = \begin{bmatrix} - \sigma_1^T - \\ - \sigma_1^T - \\ \vdots \\ - \sigma_n^T - \end{bmatrix}$, $X_i = \sigma_i^T \underline{Y}$, let $f(\underline{Y}) = \max_{i \in [n]} \{\sigma_i^T y\}$

According to Princare' theorem, if f is L-lipschitz, then we have

$$\forall \underline{y}, \underline{y}' : |f(y) - f(y')| \leq \mathcal{L} \|y - y'\|_2 \longrightarrow Var(f(y)) \leq \mathcal{L}^2$$

Y_i's are independent

let $f(Y) = \sigma_j^T Y$, $f(Y') = \sigma_k^T Y'$, also let $\sigma_j^T y \geq \sigma_k^T y'$.

$$\rightarrow |f(Y) - f(Y')| = |\sigma_j^T Y - \sigma_k^T y'| \leq |\sigma_j^T (y - y')| \leq \|\sigma_j\|_2 \cdot \|y - y'\|_2$$

$$\rightarrow |f(y) - f(Y')| \leq \sqrt{\max_j \|\sigma_j\|^2} \cdot \|y - y'\|_2 = \mathcal{L} \|y - y'\|_2$$

then the lipschitz constant would be : $\mathcal{L} = \sqrt{\max\{\sigma_1^2, \dots, \sigma_n^2\}} = \|A\|_{op}$

on the Other hand:

$\hookrightarrow \mathcal{L} = \max_{i \in [n]} \sqrt{\sigma_i^2}$

$$\mathbb{E}[X_i] = \sigma_i^T \mathbb{E} Y_i = 0$$

$$Var(X_i) = \sigma_i^T \mathbb{E}[\underbrace{YY^T}_{I}] \sigma_i = \sigma_i^T \sigma_i = \|\sigma_i\|_2^2$$

Q.E.D. ■

$$\rightarrow Var[f(y)] = Var\{\max_i \{X_i\}\} \leq \mathcal{L}^2 = \max_{i \in [n]} \{Var(X_i)\}$$

:60/5

(I) $\underline{X} = (X_1, \ldots, X_n)$, $\underline{X}_i' = (X_1, \ldots, X_i', \ldots X_n)$, $\mu = \mathbb{E} f(\underline{X})$

$$\mathrm{Var}[f(\underline{X})] \le \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[\left(f(\underline{X}) - f(\underline{X}_i')\right)^2\right] = \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left(\left((f(\underline{X}) - \mu) - (f(\underline{X}_i) - \mu)\right)^2\right)$$

$$\hookrightarrow \mathrm{Var}[f(\underline{X})] \le \mathbb{E}\left[\sum_{i=1}^{n} \mathrm{Var}_i(f(\underline{X}))\right] \times \cancel{\frac{1}{2}} \times \cancel{2}$$

On the other hand:

$$\mathrm{Var}_i[f(\underline{X})] \le \left(\frac{1}{2}\right)^2 \cdot \left[\sup_z \{f(X_{1:i-1}, z, X_{i+1:n})\} - \inf_z \{f(X_{1:i-1}, z, X_{i+1:n})\}\right]^2 = \frac{1}{4} D_i^2[f(\underline{X})]$$

thus, we get:

$$\mathrm{Var}[f(\underline{X})] \le \frac{1}{4} \sum_{i=1}^{n} D_i^2(f(\underline{X}))$$ Q.E.D. ■

(II) obviously, $D_i(f(\underline{X})) \le 1$. Thus from above: $\mathrm{Var}(f(\underline{X})) \le \frac{1}{4} \sum_{i=1}^{n} D_i^2(f(\underline{X})) \le \frac{n}{4}$

Furthermore, since $B_n = \sum_{i=1}^{n} X_i \rightarrow \mathbb{E}[B_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n \times \frac{1}{2} = \frac{n}{2}$

Therefore: $\left.\begin{array}{l} \mathrm{Var}(f(\underline{X})) \le \frac{n}{4} \\ \mathbb{E}(f(\underline{X})) \ge \frac{n}{2} \end{array}\right\}$ $\mathbb{E}[B_n] = \mathbb{E}[f(\underline{X})] \ge \frac{n}{2} + O(\sqrt{n})$

this means that the concentration is around $\frac{n}{2}$, and since the variance is bounded by $O(n)$, there is a good probability that we are around $\frac{n}{2}$.

:۹۰۱ن

① $Z_k(\theta) = X_k \sin\theta + Y_k \sin\theta \rightarrow \dot{Z}_k(\theta) = X_k \cos\theta - Y_k \sin\theta$

$\rightarrow \dot{\underline{Z}}(\theta) = \underline{X}\cos\theta - \underline{Y}\sin\theta$. Since $\underline{X}, \underline{Y}$ are independent Gaussians, we can say that they're jointly gaussians and thus their linear combination is gaussian as well. Therefore, both $\underline{Z}(\theta)$ & $\dot{\underline{Z}}(\theta)$ are gaussians.

① $\mathbb{E}[\underline{Z}(\theta)] = \mathbb{E}(\sin\theta\, \mathbb{E}(\underline{X}) + \cos\theta\, \mathbb{E}(\underline{Y})) = 0$

② $\mathbb{E}[\underline{Z}(\theta)\underline{Z}^T(\theta)] = \mathbb{E}[\sin^2\theta\, \underline{X}\underline{X}^T] + \mathbb{E}[\cos^2\theta\, \underline{Y}\underline{Y}^T] + \mathbb{E}[\underbrace{\sin\theta\cos\theta\, \underline{X}\underline{Y}^T}_{0}] + \mathbb{E}[\underbrace{\sin\theta\cos\theta\, \underline{Y}\underline{X}^T}_{0}]$

$= (\sin^2\theta + \cos^2\theta) I_n = I_n$

③ $\mathbb{E}[\dot{\underline{Z}}(\theta)] = \cos\theta\, \mathbb{E}(\underline{X}) - \sin\theta\, \mathbb{E}(\underline{Y}) = 0$

④ $\mathbb{E}[\dot{\underline{Z}}(\theta)\dot{\underline{Z}}^T(\theta)] = \mathbb{E}[\cos^2\theta\, \underline{X}\underline{X}^T] + \mathbb{E}[\sin^2\theta\, \underline{Y}\underline{Y}^T] = I_n$

⑤ $\mathbb{E}[\underline{Z}(\theta)\dot{\underline{Z}}^T(\theta)] = \mathbb{E}(\cos^2\theta\, \underline{X}\underline{X}^T) - \mathbb{E}(\sin^2\theta\, \underline{Y}\underline{Y}^T) = I_n - I_n = \emptyset_{n\times n}$

$\rightarrow \underline{Z}(\theta) \sim \mathcal{N}(0, I_n)$, $\dot{\underline{Z}}(\theta) \sim \mathcal{N}(0, I_n)$, $\underline{Z}(\theta) \perp\!\!\!\perp \dot{\underline{Z}}(\theta)$

---

② $\mathbb{E}_X[\Phi(f(X) - \mathbb{E}_X[f(X)])] = \mathbb{E}_X[\Phi(f(X) - \mathbb{E}_Y[f(Y)])] = \mathbb{E}_X(\Phi(\mathbb{E}_Y(f(X) - f(Y))))$

by Jensen (Φ is convex)

$\leq \mathbb{E}_{X,Y}[\Phi(f(X) - f(Y))] = \mathbb{E}_{X,Y}[\Phi(\int_0^{\pi/2} \frac{d}{d\theta} f(Z(\theta))\, d\theta)]$

$= \mathbb{E}_{X,Y}[\Phi(\int_0^{\pi/2} \langle \nabla f(Z(\theta)), \dot{Z}(\theta)\rangle\, d\theta)] \leq \frac{2}{\pi}\mathbb{E}_{X,Y}[\int_0^{\pi/2} \Phi(\frac{\pi}{2}\langle \nabla f(Z(\theta)), \dot{Z}(\theta)\rangle)\, d\theta]$

ادامه سوال ۲:

Since the expectation is over X, Y and not θ:

$$\mathbb{E}_{X,Y}\left[\Phi\left(\frac{\pi}{2}\langle\nabla f(Z(\theta)), \dot{Z}(\theta)\rangle\right)\right] = \mathbb{E}_{X,Y}\left[\Phi\left(\frac{\pi}{2}\langle\nabla f(Z(0)), \dot{Z}(0)\rangle\right)\right]$$

$$= \mathbb{E}_{X,Y}\left[\Phi\left(\frac{\pi}{2}\langle\nabla f(\underline{Y}), \underline{X}\rangle\right)\right] = \mathbb{E}_{X,Y}\left[\Phi\left(\pi/2\,\langle\nabla f(\underline{X}), \underline{Y}\rangle\right)\right]$$

$$\Rightarrow \mathbb{E}_X\left[\Phi\left(f(X) - \mathbb{E}f(X)\right)\right] \leq \frac{2}{\pi} \times \frac{\pi}{2}\, \mathbb{E}_{X,Y}\left[\Phi\left(\frac{\pi}{2}\langle\nabla f(\underline{X}), \underline{Y}\rangle\right)\right]$$

Q.E.D.

(III) $\Phi(X) = e^{\lambda X}$

$$\longrightarrow \mathbb{E}_X\left[e^{\lambda(f(X) - \mathbb{E}f(X))}\right] \leq \mathbb{E}_X\,\mathbb{E}_Y\left[e^{\overbrace{\frac{\lambda\pi}{2}\langle\nabla f(X), Y\rangle}^{T(Y|X)}}\,\Big|\, X\right]$$

since $\mathbb{E}\left[T(Y)\,|\,X\right] = 0$, $\mathrm{Var}\left(T(Y)\,|\,X\right) = \frac{\lambda^2\pi^2}{4}\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}(X_i)\right)^2 \mathrm{Var}(Y_i) = \frac{\lambda^2\pi^2}{4}\|\nabla f(X)\|_2^2$

→ thus: $\mathbb{E}_Y\left[e^{T(Y|X)}\right] = \exp\left[\frac{\lambda^2\pi^2}{8}\|\nabla f(\underline{X})\|_2^2\right]$

L-lipschitz: $\|\nabla f(x)\| \leq L$

$$\rightarrow \mathbb{E}_X\left[e^{\lambda(f(X) - \mathbb{E}f(X))}\right] \leq \mathbb{E}_X\left[e^{\frac{\lambda^2\pi^2}{8}\|\nabla f(X)\|_2^2}\right] \leq \exp\left(\frac{\lambda^2\pi^2 L^2}{8}\right)$$

So $\Phi(X)$ is a subgaussian Random Variable with $\sigma^2 = \left(\frac{L\pi}{2}\right)^2$. Q.E.D.

(IV) Since we used the Properties of Gaussian vectors (& the fact that only in gaussian distribution uncorrelatedness ≡ independentness). So we can't apply the same Reasoning for subgaussian vectors.

Scanned by CamScanner