HW2 High dimensional Probability Kiarash Gheisaripour

402102302

Q1) M is a δ-net for K, therefor: $\forall p \in K : \exists p_m \in M$ so that $d(p, p_m) \leqslant \delta$.

similiarly N is an ε-net for M therefore we have: $\exists p_n \in N : d(p_n, p_m) \leqslant \varepsilon$

$$\rightarrow \quad d(p, p_n) \underset{\substack{\text{(triangle} \\ \text{ineq.)}}}{\leqslant} d(p, p_m) + d(p_n, p_m) \leqslant \delta + \varepsilon$$

by the ineq. above we have shown for every $p \in K$ we have a $p_n \in N$ so that $d(p, p_n) < \delta + \varepsilon$, this Concludes the proof $\square$.

1.2) the left side ineq is trivial because: if M is an internal ε-net, it is also an external ε-net

$$\rightarrow \quad \{\text{internal}, \varepsilon\text{-nets}\} \subseteq \{\text{external } \varepsilon\text{-nets}\}$$

$$\rightarrow \quad \inf_{\substack{M:\text{int} \\ \varepsilon\text{-net}}} |M| \geqslant \inf_{\substack{M:\text{ext} \\ \varepsilon\text{-net}}} |M|$$

$$\rightarrow \quad N^{ext}(K, d, \varepsilon) \leqslant N(K, d, \varepsilon)$$

1.2) continued:

the right inequality is a tad bit more tricky

we want to show every $\frac{\varepsilon}{2}$-net$_{(ext)}$ is transformable

to an ε-net$_{(int)}$ with conserving the Cardinality

of the set, which is equivalent to

$\mathcal{N}(\mathcal{K}, d, \varepsilon) \leq \mathcal{N}^{ext}(\mathcal{K}, d, \frac{\varepsilon}{2})$ (by the same argument as the left ineq.)

say that $\mathcal{M}$ is an (ext) $\frac{\varepsilon}{2}$-net

therefore for every $p \in \mathcal{K}$ we have $p_m \in A$

: $d(p, p_m) < \frac{\varepsilon}{2}$, say that luck has it

so that $p_m \notin \mathcal{K}$, by the fact that

$\mathcal{M}$ is an (ext)($\frac{\varepsilon}{2}$-net) there exists an arbitrary

$p_{in} \in K$ so that $d(p_m, p_{in}) < \frac{\varepsilon}{2}$

$\rightarrow$ $d(p, p_{in}) \leq d(p_m, p_{in}) + d(p_m, p) \leq \varepsilon$

replace p_m with p_{in} in $\mathcal{M}$.

do this for $\{p_m \mid p_m \in \mathcal{M}, p_m \notin \mathcal{K}\}$ and

you'll have an $\mathcal{M}$ contained in $\mathcal{K}$

that concludes the proof. $\square$

(Q 1.3) to prove the right ineq., take an arbitrary optimal ε-Packing set M, we want to show that this set is also an ε-net.

we'll use proof by contradiction:

say that there exists a $p \in X$

so that: $d(p, p_m) > \varepsilon : \forall p_m \in M$, this contradicts with the fact that M is an optimal ε-packing! this Concludes the RHS ineq. proof.

to show that the left ineq. holds we'll show that the Cardinality of every 2ε-packing is smaller than every ε-net

take an arbitrary 2ε-p, ε-net: M, N

we know that for every $P_{m_i} \in M$, $\exists P_{n_j} \in N$

$\rightarrow d(P_{m_j}, P_{n_j}) < \varepsilon$ take $f: M \rightarrow N$

$\rightarrow f(P_{m_j}) = P_{n_j}$, we'll show that this function is an injective function.

if $f(P_{m_j}) = f(P_{m_{j'}}) = P_{n_j} \rightarrow d(P_{m_j}, P_{m_{j'}}) < d(P_{m_j}, P_{n_j}) + d(P_{m_{j'}}, P_{n_j})$
$< 2\varepsilon$

which Contradicts the 2ε-packing'ness $\rightarrow |M| \leq |N|$

CamScanner

Q1.4) counter example Construction:

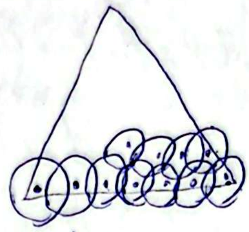
Imagine a set L which is a non-optima ε-net on K

the set L is also an $\varepsilon+\delta$-packing which means

for $\forall x,y \in L$, $d(x,y) > \varepsilon+\delta$. this is not a very

unnormal constraint as it can easily be

seen Geometrically:



(if δ is small ($\delta \ll \varepsilon$)

this is a making of a bad (non-opt) covering)

because L is a non optimal covering

then $|L| > N(K,d,\varepsilon)$. i'll argue that $N(L,d,\varepsilon)$

$= |L|$, if we want to take away any point from (P_a)

L the set itself won't be able to cover it

because $\forall p \in L/\{P_a\} : d(P, P_a) \geq \varepsilon+\delta > \varepsilon$

therefore: $N(L,d,\varepsilon) > N(K,d,\varepsilon)$

1.4) Consider an $\frac{\varepsilon}{2}$-net on K, A

define B as followed: $B := \{p : p \in L, p \in A\}$

we want to show that B covers L. (when we add some elements)

take an arbitrary point in L, p

if p is covered by a $P_c \in B$, then Good!

Q 1.4) continued:

if P_c is not in B, then there exists a P_c' in L

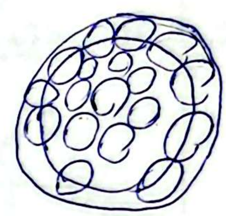
for which : $d(P_c, P_c') \leqslant \frac{\varepsilon}{2}$

therefore : $d(P, P_c') \leqslant d(P, P_c) + d(P_c', P_c) < \varepsilon$

if we add all $P_c'(s)$ to B (each for every $P_A \in A$ which is not in L then B becomes an ε-net for L , $|B| = |A|$) therefore by a similar argument to the first parts of Q1) the ineq holds.

1.5) $N(K, d, \varepsilon) \leqslant P(K, d, \varepsilon)$

we'll bound P using volume :



$P \times Vol(B^n(\varepsilon)) < Vol(B^n(1+\frac{\varepsilon}{2}))$

$\rightarrow \quad P \times (2\varepsilon)^d < (2+2\varepsilon)^d \rightarrow P < \left(\frac{1+\varepsilon}{\varepsilon}\right)^d$

this concludes the proof.

1.6) consider an m-net A on K

$\forall P \in K \ \exists P_c \in A : \ d(P, P_c) \leqslant m$

this means that by flipping at most m-bits we can turn every $P_c \in A$ into any desirable points. meaning that P_c is with $\sum_{k=0}^{m} \binom{n}{k} \rightarrow$ k flips points in radius m.

1.b) continued:

therefore by flipping all the $P_c \in A$ we will cover all $p \in \mathcal{K}$ atleast once: $2^n \leq |A| \sum_0^m \binom{n}{k}$

$$\rightarrow \frac{2^n}{\sum_0^m \binom{n}{k}} \leq N(\mathcal{K}, d_H, m)$$

now we'll show that for an optimal packing the right ineq. holds.

in order to show this we'll show that in an optimal m-packing: $\cancel{\exists p \in \mathcal{K}, \forall p_m \in A: d(p, P_m) \geq \frac{m}{2}}$

~~we proof by Contradiction. say that~~ $\cancel{\forall p \in \mathcal{K} : d(P, P_m) \leq \frac{m}{2}}$ $\cancel{\exists P_m \in A}$

if for every $P_m \in A$ we calculate the (air quote) "volume" of the Ball with radius $\lfloor \frac{m}{2} \rfloor$ which corresponds to the $\lfloor \frac{m}{2} \rfloor$ neighbourhood: $\sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} \binom{n}{k}$

now we'll show that these Balls can't quite take the whole "volume" of $\mathcal{K}$. Because this is an m-packing $d(P_{m_i}, P_{m_j}) > m$

even if we flip the bits in the "right" direction we will still have $d(P'_{m_i}, P'_{m_j}) > m - 2\lfloor \frac{m}{2} \rfloor \geq 0$

which means these balls have no overlap

1.6) continued:

Since these balls are also contained in $\mathcal{K}$ (like how are they Gonna get out?! :D) the sum of their volumes is less than the whole volume:

$$|A| \times \sum_{1}^{\lfloor \frac{m}{2} \rfloor} \binom{n}{k} \leq 2^n \rightarrow M(\mathcal{K}, d_H, m) \leq \frac{2^n}{\sum_{1}^{\lfloor m/2 \rfloor} \binom{n}{k}}$$

1.7) first off because we rescaled our metric our ineq changes to

$$M(\mathcal{K}, d_H^n, \delta) \leq \frac{2^n}{\sum_{k=1}^{\lfloor \frac{\delta n}{2} \rfloor} \binom{n}{k}}$$

from Q5 of (pset1) we know that:

$$\frac{1}{n+1} e^{-nD(\frac{\delta}{2} \| \frac{1}{2})} \leq P(Z_n \leq \lfloor \frac{\delta n}{2} \rfloor) = \sum^{\lfloor \frac{\delta n}{2} \rfloor} \binom{n}{k} \left(\frac{1}{2}\right)^n$$

$$\rightarrow M(\mathcal{K}, d_H^n, \delta) \leq (n+1) e^{n(D(\frac{\delta}{2} \| \frac{1}{2})}$$

$$\rightarrow \frac{\log(M(\mathcal{K}, d_H^n, \delta)}{d} \leq \frac{\log^{d+1}}{d} + D_{KL}(\delta_{/2} \| 1/2)$$

Q2)

2.1) the upper bound is trivial because:

$$P\left(\bigcup_1^n A_i\right) \leqslant \sum P(A_i) \quad \text{(union bound)}, \quad P\left(\bigcup_1^n A_i\right) \leq 1$$

$$\rightarrow \quad P\left(\bigcup_1^n A_i\right) \leqslant 1 \wedge \sum P(A_i)$$

lower bound: $\left(\bigcup_1^n A_i\right)^c = \bigcap_{i=1}^n A_i^c$

$$\rightarrow \quad P\left(\bigcup A_i\right) = P\left(\left(\bigcap A_i^c\right)^c\right) = 1 - P\left(\bigcap A_i^c\right)$$

$$= 1 - \prod (1 - P(A_i)) \geqslant 1 - \exp\left(\sum P(A_i)\right)$$

$$\geqslant (1-e^{-1})\left(1 \wedge \sum P(A_i)\right)$$

2.2)

$$P(\sup X_t > x) = P\left(\bigcup_{t \in T} (X_t > x)\right).$$

$$\rightarrow \quad P\left(\sup X_t \geqslant \eta^{*-1}(\log|T| + x)\right) \geqslant (1-e^{-1})\left(1 \wedge \sum P(X_t > x)\right)$$

$$\geqslant (1-e^{-1})\left(1 \wedge e^{-\eta^*(\eta^{*-1}(\log|T|+x))} \times |T|\right)$$

$$= (1-e^{-1})\, e^{-x}$$

2.3)

$$P\left(\sup X_t > \frac{1}{2}\eta^{*-1}(2\log|T| + x)\right) \geqslant (1-e^{-1}) \exp\left(-\eta^*\left(\frac{1}{2}\eta^{*-1}(2\log|T|+x)\right.\right.$$

2.3) continued: if f is Concave then $\frac{f(x)+f(y)}{2} < f\left(\frac{x+y}{2}\right)$

$$\rightarrow \quad \frac{\eta^{*-1}(2\log|T|)+\eta^{*-1}(\eta(2x))}{2} < \eta^{*-1}\left(\log|T|+\frac{\eta^{*}(2x)}{2}\right)$$

there for we have: $P[\sup X_t > \cdots] \geq (1-e^{-1})\exp\left(\frac{-\eta^{*}(2x)}{2}\right)$

2.4) first we'll show the upper bound

$$\sup X_t \leqslant \sup X_t \wedge 0 \rightarrow \mathbb{E}\sup X_t \leqslant E(\sup X_t \wedge 0)$$

$$= \int_0^\infty P(\sup X_t > u)du = \int_0^{\psi^{*-1}(\log|T|)} P(\cdots) + \int_{\psi^{*-1}(\log|T|)}^\infty P(\cdots)$$

$$\leqslant \underbrace{\psi^{*-1}(\log|T|)}_{\text{call this } \beta} + \int_\beta^\infty P(\sup X_t > u)\,du \leqslant \beta + |T|\int_\beta^\infty e^{-\psi^{*}(u)}du$$

(union bound)

note that $\psi^{*}(x)$ is Convex, therefore we have:

$$\psi^{*}(x) > \psi^{*}(\beta) + (x-b)\,\psi^{*\prime}(\beta)$$

$$\rightarrow \beta + |T|\int_\beta^\infty e^{-\psi^{*}(u)}du \leqslant \beta + \int_0^\infty e^{-x\psi^{*\prime}(\beta)}dx = \beta + \frac{1}{\psi^{*\prime}(\beta)}$$

strictly decreasing term

$$= O(\beta) = C_2\,\psi^{*-1}(\log|T|)$$

Q2.4) continued:

now we prove the lower bound:

Note that: $(\sup X_t) = (\sup X_t \wedge 0 + (\sup X_t)\vee 0$ therefore

$$\mathbb{E} \sup X_t = E \sup X_t\wedge 0 + E \sup X_t \vee 0$$

$$\geq E \sup X_t \vee 0 + \sup E X_{t\wedge 0}$$
(jensen)

$\Big(E \sup X_t \vee 0 = \int_0^\infty P(\sup X_t \geq a) du \geq \int_0^a P(Y \geq a) du \geq a e^{-a}(1-e^{-1})$

$a = \eta^{*-1}(\lg |T|)$

$\to E \sup X_t \geq 1-e^{-1} \eta^{*-1}(\lg |T|) + \sup E X_{t\wedge 0}$

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

$$\geq \frac{1}{\sqrt{2\pi}} e^{-x^2} \int_0^\infty e^{-v^2} dv = \frac{e^{-x^2}}{2\sqrt{2}}$$

2.6) $e^{-\eta^*(x)} \leq P(X \geq x) \leq e^{-\psi^*(x)}$

$\psi^*(a) = \frac{a^2}{2} \to \psi^{*-1}(x) = \sqrt{2x}$

$\eta^{*}(x) = \frac{a^2}{2} + \lg 2\sqrt{2} \to \eta^{*-1}(x) = \sqrt{2(x - \lg 2\sqrt{2})}$

2.6) continued:

lower bound: $(1-e^{-1})\ \eta^{*^{-1}}(\lg n) + \sup E\ on\ X_t$

$$= 1-e^{-1}\sqrt{2\lg(n\,2^{-3/2})} + \int_{-\infty}^{0} \frac{x\,e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx$$

$$= \frac{1-e^{-1}}{2}\sqrt{2\lg(n\,2^{-3/4})} - \frac{1}{\sqrt{2\pi}}$$

upper bound: with our ineq. from befor

it's obvious $E\sup X_t \leqslant O(\sqrt{2\lg n})$

but to show it explicitly:

$$\exp(\lambda E\max Z_i) \leqslant E\exp(\lambda\max Z_i)$$

$$\leqslant N\,E e^{\lambda Z_i} = n e^{\frac{\lambda^2}{2}} \rightarrow \text{minimize w.r.t } \lambda$$

$$\rightarrow E\sup X_t = \frac{\lg N}{\lambda} + \frac{\lambda}{2}$$

$$\rightarrow E\sup X_t \leqslant \sqrt{2\lg N}$$

Q3) in order to prove the given norm is bounded like so w.h.p., we'll first write the operator norm in the supremum form and try to benefit from Quantization method:

$$A \in \mathbb{R}^{n\times n} \qquad \|A\|_{op} = \sup_{\substack{x\in \\ \|x\|=1}} x^T A x$$

$$; \qquad \|x^T A x - y^T A y\| \leqslant \|x^T A (x-y)\| + \|(x-y)^T A y\| \leqslant 2\|A\|\,\|x-y\|$$

therefore by introducing an ε-net N we'll have

$$\|A\| = \sup_{\|x\|=1} x^T A x \leqslant 2\varepsilon \|A\| + \sup_{x\in N} x^T A x$$

$$\rightarrow \quad \|A\| \leqslant \frac{\sup_{x\in N} x^T A x}{1-2\varepsilon} \longrightarrow$$ if we bound this we've bound $\|A\|$

$$\left\|\frac{1}{m}A^T A - I_n\right\|_{op} = \sup_{\|x\|=1} \frac{x^T A^T A x}{m} - 1$$

$$x^T A^T A x = \|Ax\|_2^2 \quad , \quad A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{pmatrix} \rightarrow Ax = \begin{pmatrix} \langle A_1, x\rangle \\ \langle A_2, x\rangle \\ \vdots \\ \langle A_n, x\rangle \end{pmatrix}$$

$$\langle A_i, x\rangle = \sum A_{ij} x_j \sim \text{subg}\left(\max_j \|A_{ij}\|_{\psi_2}\right)$$

$$\|Ax\|_2^2 = \sum_i \langle A_i, x\rangle^2 \longrightarrow$$ subg $\times$ subg = subexponential

Q3) continued:

because of isotropicity we have $\mathbb{E} A_j^T A_j = I$

$\to$ $\mathbb{E}\langle A_j, x\rangle^2 = 1$ $\to$ $\sup \frac{x^T A^T A x}{m} - 1 = \sup \left| \frac{1}{m} \sum (\langle A_j, x\rangle^2 - 1)\right.$

$\langle A_j, x\rangle^2 - 1 = Z_j(x)$ $\to$ Centered Sub-exp R.V.s

$\| Z_j(x)\|_{\psi_1} \leqslant C \|\langle A_j, x\rangle\|_{\psi_2}^2$ (from the Centering argument and subg x subg = sub exp)

now we apply the Bernstein inequality

$$P\left\{ \left|\sum Z_j(x)\right| \geqslant t \right\} \leqslant 2\exp\left(-c\min\left(\frac{t^2}{\sum \|X_j\|_{\psi_1}^2}, \frac{t}{\max_j \|X_j\|_{\psi_1}}\right)\right)$$

now we want to use the arguements stated above to conclude our original Statement

$N(B_2^n, \varepsilon) \leqslant \left(\frac{2}{\varepsilon}+1\right)^n$ $\to$ if we set $\varepsilon = \frac{1}{4}$

$|N| \leqslant 9^n$, $\|\underbrace{\frac{A^T A}{m} - I}_{B}\|_{op} \leqslant 2 \sup_{x \in N} x^T B x$

therefore if we bound $E := \{ \sup x^T B x \leqslant C' \max\{\delta, \delta^2\}\}$:

we have essentially bound $P(\|B\|_{op} \leqslant \cdots)$

Q3) continued:

$$P\left(\left|\frac{1}{m}\sum(Z_i(n)^2-1)\right|\geq\frac{\varepsilon}{2}\right)\leq 2\exp\left(-c_1\min\left(\frac{\varepsilon^2}{K^4},\frac{\varepsilon}{K^2}\right)m\right)$$

$$\frac{\varepsilon}{K^2}=\max(\delta,\delta^2)\quad\rightarrow\quad=2\exp(-c_1\delta^2 m)\leq 2\exp(-c_1C^2(n+t^2))$$

now using the union bound we have:

$$P\left(\max_{n\in N}\left|\frac{1}{m}\|A_n\|_2^2-1\right|\geq\frac{\varepsilon}{2}\right)\leq 9^n\, 2\exp(-c_1C^2(n+t^2))\leq 2\exp(t^2)$$

the above argument follows if we choose the absolute Constant C big enough.

this concludes the proof.

Q4) Poincare inequality:

$$f:\mathbb{R}^d\to\mathbb{R},\quad X\sim N(0,I_d)$$

$$Var(f(X))\leq E\left[\|\nabla f(X)\|_2^2\right]$$

it follows from the above inequality that if f is L-lipschitz then $Var(f(X))\leq L^2$

going back to our Original Problem we had:

$$Var\left[\max X_i\right]\leq\max Var(X_i)$$

CamScanner

Q 4) continued:

$$X \sim N(\cdot, C) \rightarrow X = C^{1/2} Y \; , \; Y = N(\cdot, I)$$

$$C^{1/2} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \rightarrow X_i = \sum_j C_{ij} Y_j$$

$$\mathrm{Var}\, X_i = \mathrm{Var} \sum C_{ij} Y_j = \|C_i\|_2^2$$

$$\rightarrow \max_i \mathrm{Var}\, X_i = \|C_l\|_2^2 \quad \left(l = \mathrm{argmax}\ \mathrm{Var}\, X_i \right)$$

now we'll show that $\max(x)$ is an L-lipschitz

function, $f = \max(\vec{y}) = \max_i \sum C_{ij} y_j$

$$\rightarrow |f(x) - f(y)| = \max_i \sum C_{ij} x_j - \max \sum C_{ij} y_j$$

$$\left(l' = \mathrm{argmax} \left(\sum C_{ij} x_j \right) \right) \leq \sum C_{lj} (x_j - y_j) = \langle C_l, x - y \rangle$$

$$(\text{Cauchy-Schwarz}) \rightarrow \langle C_l, x - y \rangle \leq \|C_l\|_2 \, \|x - y\|_2 \leq \|C_l\|_2 \, \|x - y\|_2$$

so f is : $\|C_l\|_2$ - lipschitz

therefore $\mathrm{Var}[\max X_i] \leq \|C_l\|_2^2 = \max \mathrm{Var}\, X_i$

$\square$.

CS CamScanner

Q5)

in Question 2.4 we showed that

if $P(X_t > x) \leqslant \exp(-\psi^*(x))$

then $E \sup X_t \leqslant C \psi^{*-1}(\log |T|)$

now in a subexponential setting we have:

$X \sim \text{sub-exp}(\sigma^2, \alpha)$

$$P(X-\mu > t) \leqslant \begin{cases} e^{-\frac{t^2}{2\sigma^2}} & 0 < t \leqslant \frac{\sigma^2}{\alpha} \\ e^{-\frac{t}{2\alpha}} & t > \frac{\sigma^2}{\alpha} \end{cases}$$

$$\psi_1^*(t) = \frac{t^2}{2\sigma^2}, \quad \psi_2^*(t) = \frac{t}{2\alpha}$$

$$\rightarrow \psi_1^{*^{-1}}(t') = \sqrt{2\sigma^2 t'}, \quad \psi_2^{*^{-1}}(t') = 2\alpha t'$$

therefore we have: $\mathbb{E} \max X_i \leqslant \max\left(C_1 \psi_1^{*^{-1}}(\log n), C_2 \psi_2^{*^{-1}}(\log n)\right)$

$$\leqslant C_3 \left(\alpha \log n + \sigma \sqrt{\log n}\right) \quad \square.$$

(we wrapped the unwanted constants in C_3)

Q6) in order to show that $\left\{ \frac{G_1}{\sqrt{n}}, \dots, \frac{G_n}{\sqrt{n}} \right\} \cap R B_2^n$

is in fact an ε-net (w.h.p)

we have to show that: $\forall x \in R B_2^n \ \exists \frac{G_i}{\sqrt{n}} \in N$

so that : $P(\| x - \frac{G_i}{\sqrt{n}} \| > \varepsilon) < \delta$ (a ting decreasing amount)

pdf for $\frac{G}{\sqrt{n}}$ $\quad f(\|x\|) = \left(\frac{n}{2\pi}\right)^{n/2} \exp\left(\frac{-n}{2} \|x\|^2\right)$

$$P\left(\| x - \frac{G_i}{\sqrt{n}} \| \leq \varepsilon\right) = \int_{B_2^n(x,\varepsilon)} f(\vec{x}') \, d\vec{x}' \geq vol\left(B_2^n(x,\varepsilon)\right) \min_{x' \in B_2^n(x,\varepsilon)} f(x')$$

$$\| x - \frac{G_i}{\sqrt{n}} \| < \varepsilon \rightarrow \frac{\|G_i\|}{\sqrt{n}} \leq R + \varepsilon \rightarrow \min f(x') = \left(\frac{n}{2\pi}\right)^{n/2} \exp\left(\frac{-n}{2}(R+\varepsilon)^2\right)$$

$$\rightarrow P\left(\| x - \frac{G_i}{\sqrt{n}} \| \leq \varepsilon\right) \geq C \varepsilon^n \left(\frac{n}{2\pi}\right)^{n/2} \exp\left(\frac{-n}{2}(R+\varepsilon)^2\right) =$$

$$C \exp\left(n\left(\log \varepsilon + \frac{\log n}{2} - \frac{(R+\varepsilon)^2}{2}\right)\right)$$

the term $\log n$ isn't anything to worry about, if we use a better approximation of $vol(B_2^n(x,\varepsilon))$ it goes away and also it's negligible comparing to n.

so we've obtained a lower bound in the form of

$\exp(-C(R,\varepsilon) n) = P_G$

Q6) now to show that our $\mathcal{N}$ is an ε-net (w.h.p)

we'll show that our points fall into an $\frac{\varepsilon}{2}$-net

(w.h.p) then we've an ε-net, because if $\mathcal{M}$ is an

$\frac{\varepsilon}{2}$-net then $\to$ $d\left(x, \frac{G_i}{\sqrt{n}}\right) \leqslant d(x, P_m) + d\left(P_m, \frac{G_i}{\sqrt{n}}\right) \leqslant 2 \times \frac{\varepsilon}{2} = \varepsilon$

$\forall x \in (R\, B_2^n)$

fix an $P_m \in \mathcal{M}$, $\forall \frac{G_i}{\sqrt{n}} \in \mathcal{N}$, $\mathbb{P}\left(\left\|\frac{G_i}{\sqrt{n}} - P_m\right\| > \varepsilon\right) \leqslant (1 - P_G)^m$

$\leqslant \exp(-m P_G)$ $\quad \left(|\mathcal{M}| \leqslant \left(\frac{CR}{\varepsilon}\right)^n \right)$

now we'll do a union Bound for $\forall P_m \in \mathcal{M}$

$$\mathbb{P}\left(\bigcup_{\substack{P_m \in \mathcal{M} \\ \frac{G_i}{\sqrt{n}} \in \mathcal{N}}} \left\|\frac{G_i}{\sqrt{n}} - P_m\right\| > \varepsilon\right) \leqslant |\mathcal{M}| \exp(-m P_G) \leqslant \left(\frac{C'R}{\varepsilon}\right)^n \exp(-m P_G)$$

$$= \exp\left(n \log\left(\frac{C'R}{\varepsilon}\right) - m \exp(-Cn)\right)$$

if we choose $m \geqslant e^{C'n}$ with C' large enough

to dominate C we can make this bound to be

$\leqslant \exp(-cn) \to \mathbb{P}\left(\left\{\frac{G_1}{\sqrt{n}}, \dots, \frac{G_m}{\sqrt{n}}\right\} \cap R B_2^n \text{ is an } \varepsilon\text{-net}\right)$

is greater than $1 - e^{-cn}$ $\square$.

Q7) from Mc-diarmid inequality

it follows that any function with the bounded difference property is sub-gaussian $\left(\frac{\sum c_i^2}{4}\right)$

we have Efron Stein inequality in this form:

$$Var(Z) \leq \mathbb{E}\left[\sum Var(Z \mid X_{\sim i})\right]$$

the random variable $Z \mid X_{\sim i}$ has the bounded difference property (c_i)

therefore is subgaussian $\left(\frac{c_i^2}{4}\right) \to Var(Z \mid X_{\sim i}) \leq \frac{c_i^2}{4}$

$$\to Var(Z) \leq \frac{\sum c_i^2}{4} \quad \square.$$

7.2) it's obvious that in this setting:

$D_i(f) \leq 1$ (because at most extreme 1 bin will be given to i or none)

$$\to Var(B_n) \leq \frac{1}{4} \sum D_i^2(f) = \frac{n}{4}$$

$$\sum X_i \leq B_n \to \mathbb{E} \sum X_i \leq \mathbb{E} B_n \to \frac{n}{2} \leq \mathbb{E} B_n$$

(Best Config everything fits (impossible almost))

Q7) continued:

since we have an upper bound on $Var\, B_n$ and a lower bound on $E B_n$ we can use chebyshev to get a Concentration Statement:

$$P\left(B_n \leq \frac{n}{2}-t\right) \leq P\left(B_n \leq E[B_n]-t\right) \leq P\left(|B_n - EB_n| \geq t\right)$$

$(E B_n \geq n/2)$

$$\leq \frac{Var\, B_n}{t^2} \leq \frac{n}{4t^2}$$

this means that that B_n is asymptotically at least $\frac{n}{2}$.

Q8)

first note that: $B = \sum_1^n X_i \quad X_i \sim Ber(p)$

$$f(B) = f(\sum X_i) = g(X_{1:n})$$

Efron - Stein inequality States that:

$$Z = g(X_{1:n}) \quad Z_i' = g(X_{\sim i}, X_i')$$

$$Var\, Z \leqslant \frac{1}{2} \sum \mathop{E}_{\substack{X_{1:n} \\ X'_{1:n}}} (Z - Z_i')^2$$

lets Calculate $\mathop{\mathbb{E}}_{X_j, X_j'} (Z - Z_j')^2$.

define S_i : $B = S_i + X_i \quad , B' = S_i + X_i'$

law of total Expectation

$$P(X_i = X_{j'}) = p^2 + (1-p)^2 \rightarrow (Z - Z_i')^2 = 0$$

$$P(X_i \neq X_i') = 2p(1-p) \Rightarrow E(f(S_i+1) - f(S_i))^2$$

$$\rightarrow E_{X_i, X_i'}[Z - Z_i'] = E_{S_i}[(f(S_i+1) - f(S_i))^2] \times 2p(1-p)$$

$$\sum E(Z - Z_i')^2 = \mathop{E}_{X_{1:N}}[\sum (f(S_i+1) - f(S_i))^2]$$

since $\#\{X_i = 1\} = B$, $\#\{X_i = 0\} = n - B$

Q8.2)

from 8.1 we have:

$$\mathrm{Var}(f(B_n)) \leq \mu\, \mathbb{E}\left[\left(1-\frac{B_n}{n}\right)(f(B_n+1)-f(B_n))^2\right]$$

$f(B_n) \to f(X)$ in dist. $\to$ $\mathrm{var}\, f(B_n) \to f(X)$

$1-\frac{B_n}{n} \to 1$ in probability

(the above is true because $P(S_n > \sqrt{n}) \leq e^{-\mu}\left(\frac{e\mu}{\sqrt{n}}\right)^{\sqrt{n}}$)

and we have $(f(B_n+1)-f(B_n))^2 \to (f(X+1)-f(X))^2$ in dist.

and since $\sup |f(K+1)-f(K)| < \infty$

therefore $\left(1-\frac{B_n}{n}\right)(f(B_n+1)-f(B_n))^2$ is bounded

therefore we have $\lim_{n\to\infty} E\left[\left(1-\frac{B_n}{n}\right)(f(B_n+1)-f(B_n))^2\right]$

$= E\left[(f(X+1)-f(X))^2\right]$ $\square$.

CamScanner

Q8) continued:

therefore $\frac{1}{2} E_{X_1:X_n}\left[\sum (f(S_i+1) - f(S_i))^2\right]$

$$= p(1-p)\, E\left[B(f(B)-f(B-1))^2 + (n-B)(f(B+1)-f(B))^2\right]$$

$$\left(X_i = 1 \rightarrow f(S_i+1) - f(S_i) = f(B) - f(B-1) \right) \cdot \#(B)$$

$$\left(X_i = 0 \rightarrow f(S_i+1) - f(S_i) = f(B+1) - f(B) \right) \#(n-B)$$

now we'll show that:

$$p(1-p)\, E\left[B(f(B)-f(B-1))^2 + (n-B)(f(B+1)-f(B))^2\right]$$

$$= p\, E\left[n-B\,(f(B+1)-f(B))^2\right]$$

$$\Rightarrow \underbrace{p(1-(1-p))\, E[(n-B)(f(B+1)-f(B))^2]}_{*_1} = \underbrace{p(1-p)\, E[B(f(B)-f(B-1))^2]}_{*_2}$$

$$*_1 = \sum_{0}^{n-1} P(B=K)(p)(n-K)(f(K+1)-f(K))^2$$

$$*_2 = \sum_{1}^{n} P(B=K)(1-p)K(f(K-1)-f(K))^2 \rightarrow (K = K+1)$$

$$= \sum_{0}^{n-1} P(B=K+1)(1-p)(K+1)(f(K+1)-f(K))^2$$

$$\rightarrow P(B=K+1)(1-p)(K+1) = P(B=K)(n-K)p$$

$$\left(\text{because: } \frac{P(B=K+1)}{P(B=K)} = \frac{\binom{n}{K+1}}{\binom{n}{K}} \frac{p^{K+1}(1-p)^{n-K-1}}{p^K (1-p)^{n-K}} = \frac{p}{1-p}\frac{n-K}{K+1}\right) \square.$$

Q8.3) $(\sqrt{x+1}-\sqrt{x})^2 = \dfrac{1}{(\sqrt{x+1}+\sqrt{x})^2} = \dfrac{1}{2x+1+2\sqrt{x(x+1)}}$

$\rightarrow$ we want to show that $(\sqrt{x+1}-\sqrt{x})^2 \leqslant \dfrac{1}{4x+1}$

$\rightarrow \dfrac{1}{2x+1+2\sqrt{x(x+1)}} \leqslant \dfrac{1}{4x+1} \rightarrow 2x \leqslant \sqrt{x(x+1)} \times 2$

$\Downarrow$

obvious because $x^2 \leqslant x(x+1)$

therefore we have

$$Var(Y) \leqslant E(\sqrt{x+1}-\sqrt{x})^2 \leqslant E\left[\frac{1}{4x+1}\right]$$

Q9) in order to prove this inequality

we'll first prove a lemma:

$\psi(x)$ cvx. , $\psi(0) = \psi'(0) = 0$

fenchel - legendre transform: $\psi^*(t) = \sup_\lambda (\lambda t - \psi(\lambda))$

therefore ψ^* is non-neg cvx, non-decreasing function

and

$$\psi^{*-1}(y) = \inf\{t \geq 0 : \psi^*(t) > y\}$$

we want to show that the above definition is equal

with this:

$$\psi^{*-1}(y) = \inf_\lambda \left\{ \frac{y + \psi(\lambda)}{\lambda} \right\}$$

proof:

define: $u = \inf_\lambda \frac{y + \psi(\lambda)}{\lambda}$

$\forall t \geq 0$, $u \geq t$ iff $\frac{y + \psi(\lambda)}{\lambda} \geq t \quad \forall \lambda > 0$

the latter ineq. implies that $y \geq \psi^*(t)$

therefore we have $\{t \geq 0 : \psi^*(t) > y\} = (u, \infty)$

$\rightarrow \inf\{\dots\} = u = \psi^{*-1}(y)$

Q9) continued:

now let's prove the desired inequality:

$$\exp(\lambda E \sup X_t) \underset{\text{(jensen)}}{\overset{\text{for part 2}\ \ast}{\leq}} E\exp(\lambda \sup X_t) = E \sup \exp \lambda X_t$$

$$\leq \sum E\exp \lambda X_t = |T|\, E\exp(\lambda X_t) \leq |T|\, e^{\psi(\lambda)}$$

$$\rightarrow E \sup X_t \leq \frac{\log|T| + \psi(\lambda)}{\lambda}$$

$$\rightarrow E \sup X_t \leq \inf \frac{\log|T| + \psi(\lambda)}{\lambda} = \psi^{*-1}(\log|T|)$$

9.2) take $\sup_{t\in T} X_t = Z$, $\psi^{*-1}(\log|T| + u) = f(u)$

$$P(Z > f(u)) = P(e^{sZ} \geq e^{sf(u)}) \underset{\text{markov}}{\leq} (Ee^{sZ})\, e^{-sf(u)}$$

$$\underset{\text{from } \ast \text{ in part 1}}{\leq} |T|\, e^{\psi(s)} e^{-sf(u)} \rightarrow P(Z \geq f(u)) \leq \inf_s |T|\, e^{\psi(s) - sf(u)}$$

$$= |T| \exp([\inf \psi(s) - sf(u)]) = |T| \exp(-[\sup_s sf(u) - \psi(s)])$$

$$= |T| \exp(-\psi^{*}(f(u))) = |T| \exp(-\log|T| - u) = e^{-u}$$

$\square$.

Q1)

the set T is the Collection of s-sparse vectors and we want to show that

$$G(T) = \mathbb{E}\left[\sup_{x \in T} |x^T Z|\right] = \mathbb{E}\left[\max_{|S|=s} \|Z_S\|_2\right]$$

note that $x^T Z = \sum_{1}^{s'} x_i Z_i \;,\; s' < s$

note that: $\sup_{|S|=s}(\cdot) = \sup_{|S| \leq s}(\cdot)$

so restricting to only s-sparse vectors doesn't change the supremum.

since we have $\binom{d}{s}$ different supports we can rewrite the sup in this way:

$$\sup_{x \in T} |x^T Z| = \max_{\substack{S \subset \{1,\dots,d\} \\ |S|=s}} \left(\sup_{\substack{x_S \in R^s \\ \|x_S\|_2 \leq 1}} |x_S^T Z_S| \right)$$

$\rightarrow$ Variational form of L_2-norm

$$= \max_{|S|=s} \|Z_S\|_2 \quad \square.$$

---

b) in order to show this, we'll use a well known inequality as

Q 10) b)

as it's written in Chapter 5 (r. Versh.)

if $g \sim (N(0, I_d))$ and f is 1-Lipschitz

$$P(f(g) \geq \mathbb{E} f(g) + \delta) \leq e^{-\delta^2/2}$$

first we'll show that $\|\cdot\|$ is 1-lipschitz which is obvious: $\big|\|x\|_2 - \|y\|_2\big| \leq \|x-y\|$ (triangle inequality)

we also have

$$\mathbb{E}\|w_s\|_2 \underset{\text{jensen}}{\leq} \sqrt{\mathbb{E}\|w_s\|_2^2} = \Big(\sum^{s} \mathbb{E}X_i^2\Big)^{1/2} = \sqrt{s}$$

therefore we have $P(\|w_s\|_2 \geq \sqrt{s} + \delta) \leq$

$$P(\|w_s\|_2 \geq \mathbb{E}\|w_s\|_2 + \delta) \leq e^{-\delta^2/2} \quad \square.$$

a simple union bound will extend this result too:

$$P\Big(\max_{|S|=s} \|w_s\|_2 > \sqrt{s} + \delta\Big) \leq \binom{d}{s} e^{-\frac{\delta^2}{2}}$$

$$\binom{d}{s} \leq \left(\frac{ed}{s}\right)^s \rightarrow \binom{d}{s} e^{-\delta^2} \leq \exp\left(L - \frac{\delta^2}{2}\right) \quad \left(L = s\log\frac{ed}{s}\right)$$

now we'd use the Identity $\mathbb{E}X = \int_0^\infty P(X > t)\,dt$

Q10) continued:

$$G(T_{(1)}) = E\left[\max \|w_s\|_2\right] = \int_0^{\infty} P(\max \|w_s\|_2 > t)\, dt$$

$$\leq \sqrt{s} + \int_0^{\infty} P(\max \|w_s\|_2 > \sqrt{s} + \delta)\, d\delta$$

$L = s \log \frac{ed}{s}$

$$\leq \sqrt{s} + \int_0^{\infty} \min\left\{1, \exp\left(L - \frac{\delta^2}{2}\right)\right\} d\delta \qquad \left(\delta_0 = \sqrt{2L}\right)$$

$$= \sqrt{s} + \int_0^{\delta_0} 1\, d\delta + \underbrace{\int_{\delta_0}^{\infty} e^{L - \delta^2/2}\, d\delta}_{\leq \frac{1}{\delta_0}} \leq \sqrt{s} + \delta_0 + \frac{1}{\delta_0}$$

$$\leq C\sqrt{L} = c\sqrt{s \log \frac{ed}{s}}$$

Q11)

$$Z_K(\theta) = X_K \sin\theta + Y_K \cos\theta$$

$$Z_K'(\theta) = X_K \cos\theta - Y_K \sin\theta \rightarrow \begin{pmatrix} Z_{K\theta} \\ Z_{K\theta}' \end{pmatrix} = \underbrace{\begin{pmatrix} \sin\theta & \cos\theta \\ \cos\theta & -\sin\theta \end{pmatrix}}_{A} \begin{pmatrix} X_K \\ Y_K \end{pmatrix}$$

since X_K, Y_K are jointly Gaussian then every linear Combination of them is also jointly Gaussian.

let's Calculate their Covariance matrix

$$E\, Z^T Z = \mathbb{E}\, A^T \underbrace{X^T X}_{I} A = \mathbb{E}\, A^T A = \begin{pmatrix} \sin^2+\cos^2 & \sin\theta\cos\theta - \sin\theta\cos\theta \\ \cos\theta\sin\theta - \sin\theta\cos\theta & \sin^2+\cos^2 \end{pmatrix}$$

$= I \rightarrow$ therefore $E[Z_K(\theta)\, Z_K'(\theta)] = 0$ (indepence)

11.2) first we'll show that:

$$\mathbb{E}[\phi(f(X) - \mathbb{E} f(X)] \leq E[\phi(f(X) - f(Y))]$$

$(X \sim Y)$

$$\rightarrow f(X) - \mathbb{E} f(X) = E[f(X) - f(Y) \mid X]$$

Jensen: $\phi E[f(X) - f(Y) \mid X] \leq E[\phi(f(X) - f(Y)) \mid X]$

$$\rightarrow E[\phi(fX - Ef(X)] \leq E[\varphi(f(X) - f(Y)]$$

11.2) continued:

define $h(\theta)$, a as followed: $h(\theta) := \frac{d}{d\theta} f(Z(\theta))$

$, a = \frac{\pi}{2}$

since φ is convex: $\varphi\left(\int_0^a h(\theta) d\theta\right) = \varphi\left(a \frac{1}{a}\int_0^a h(\theta) d\theta\right)$

$\leq \frac{1}{a}\int_0^a \varphi(a h(\theta)) d\theta$ (Jensen)

therefore $\varphi(fX - fY) \leq \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \varphi\left(\frac{\pi}{2} \frac{d}{d\theta} f(Z(\theta)\right) d\theta$

$\rightarrow$ $E\,\varphi(fX - fY) \leq \frac{2}{\pi}\int_0^{\frac{\pi}{2}} E\left[\varphi\left(\frac{\pi}{2}\frac{d}{d\theta} f(Z_\theta) d\theta\right.\right.$

$\frac{d}{d\theta} f(Z_\theta) = \nabla f(Z(\theta)) Z'(\theta)$

therefore since Z, Z' have same dist as X, Y, (we proved it in part a) (I cov)

$\frac{2}{\pi}\int_0^{\frac{\pi}{2}} E\left[\varphi \frac{\pi}{2}\frac{d}{d\theta} f Z(\theta)\right) d\theta = E\,\varphi\left(\frac{\pi}{2} Y^T \nabla f(X)\right)$

$\boxtimes$.

Q17) c)

for $\varphi(x) = \exp(\lambda x)$ we have:

$$\mathbb{E}\left[e^{\lambda(fX - \mathbb{E}fX)}\right] \leq \mathbb{E}\left[\exp \lambda \frac{\pi}{2} Y^T \nabla f(X)\right]$$

in order to prove $f - \mathbb{E}f$ is subgaussian we'll condition on X. note that for $v \in \mathbb{R}^n$

we have: $Y^T v \sim N(0, \|v\|_2^2)$

therefore $\mathbb{E}[\exp(tY^T v) \mid X] = \exp\left(\frac{t^2}{2}\|v\|_2^2\right)$

if we set $t = \frac{\lambda\pi}{2}$, $v = \nabla f(X)$

and note that because f is L-lip

therefor $\|\nabla f(X)\|_2^2 \leq L^2$

$$\mathbb{E}\left[\exp\left(\lambda \frac{\pi}{2} Y^T \nabla f(X)\right) \mid X\right] \leq \exp\left(\frac{\lambda^2 \pi^2 L^2}{8}\right)$$

taking $E_X[\cdot]$ from both sides will yield the wanted result.

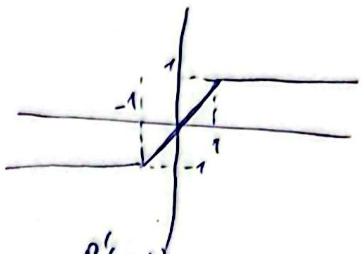
therefore we have shown: $E e^{\lambda(f - Ef)} \leq \exp \frac{\lambda^2 \pi^2 L^2}{8}$

$\rightarrow$ $~~f(X)~~ - Efx \sim \text{subgaussian}\left(\frac{\pi L}{2}\right)$

Q11) d) this inequality does not hold for subg variables in general

we'll do a Counter example:

$X \sim \text{Rademacher}(\frac{1}{2})$, $f(x)$:



$f'(\pm 1) = 0$

$\mathbb{E} f(X) = \frac{1}{2} \times 1 + \frac{-1}{2} \times 1 = 0$, therefor $Ef'(X) = 0$

$\varphi(x) = x^2$ since $f'(X) = 0$ a.s.

therefore $E[\varphi(f - Ef)] = E f^2(X) = 1$

, $Y f'(X) = 0$ a.s. because $f'(X) = 0$ a.s.

therefore $\mathbb{E}[\varphi(fX - \mathbb{E}fX)] \leq E[\varphi(\alpha Y f'(X))]$

does NOT hold.