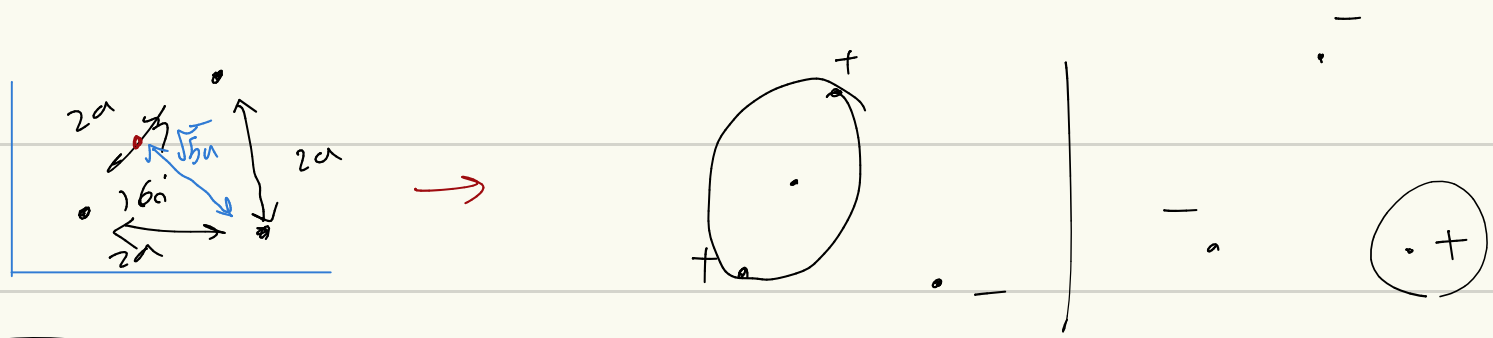
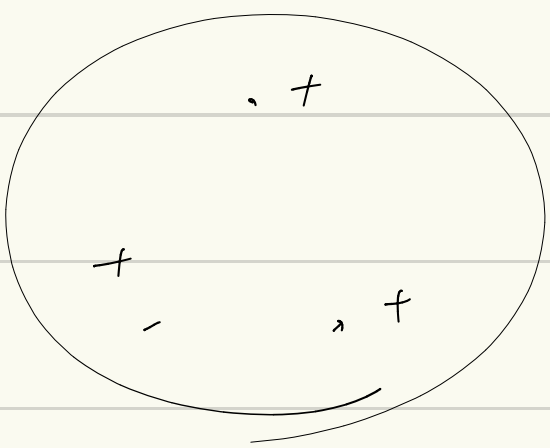
Q1)

a) first not that we will proof for a subset consist of 3 point circles in the 2D space can shatter

$$\mathcal{F} = \{ (x-a)^2 + (y-b)^2 \leq t \mid a, b, t \in \mathbb{R}, \mathbb{R}, \mathbb{R}^+ \cup \{0\} \}$$

so now consider set below

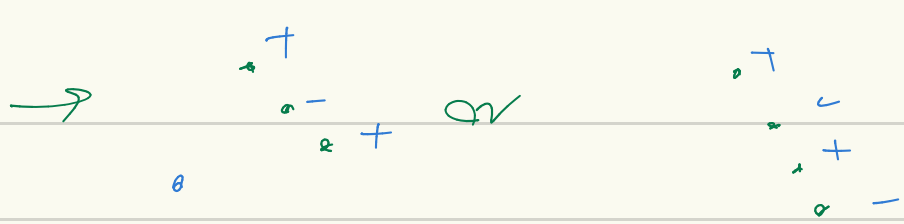




so every state of points and their labeling is classified $\Rightarrow$ $VC \geq 3$

now we proof that for any 4- arbitary point in the space shattering is not possible.

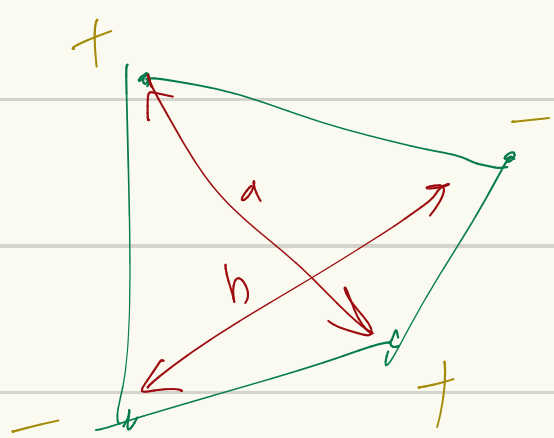
if 4 point or 3 of them are in the same line



$\rightarrow$ no circle can classify them ☹

so the another one is when 4 Points create 4-Edge Polygon. $\Rightarrow$ any arbitary Polygon in the place.

so create a convex Polygon with four Points and calculate distance between non neighbor Points. which are 2 shortline. Pick the longest and label to Node + $\Rightarrow$ there is no circle that can shatter this. $\Rightarrow$ there exist a labeling for any 4 Points that cicles can not shatter them.



$a > b$

$\Longrightarrow$ no circle to classify this!!

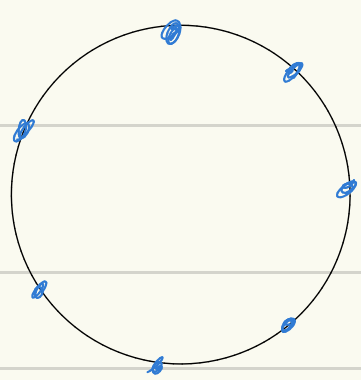
so $VC \geq 3$ and $VC < 4 \Rightarrow \boxed{VC = 3}$

b) first consider seven point on a circle

1- Point + $\rightarrow$

2 Point + $\rightarrow$ center

3 Point + $\rightarrow$



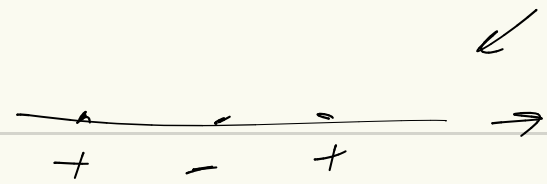
4 $\leq$ Point + $\rightarrow$ change + and −

so these seven Points are shattered by triangular

$\Rightarrow \quad VC \geq 7$

now consider any arbitrary 8 Point $\Rightarrow$

if any trille of them are in one line



$\rightarrow$ no triangular can shatter this.

so they are Node of 8-Edges Polygen

so for any 8-Point in the space create their convex Polygen $\rightarrow$ $x_1 \; x_2 \; x_3 \; x_4 \; x_5 \; x_6 \; x_7 \; x_8$

is order of the lines $\rightarrow$ x_1 neighbor of x_2, x_8

x_2 " " x_3, x_1

$\vdots$

now $x_1, x_3, x_5, x_7 \in \{-\}$

$x_2, x_4, x_6, x_8 \in \{+\}$

NO triangular can shatter this ✓

Q2) we use rademacher random variable $\varepsilon \to \begin{cases} 1 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{cases}$

So we would have.

$$P\left[\sum x_i > t\sqrt{\sum x_i^2}\right] \Rightarrow P\left[\frac{\sum x_i}{\sqrt{\sum x_i^2}} > t\right] \Rightarrow$$

$$P\left[\frac{\sum x_i}{\|x\|_2} > t\right] \overset{\text{chernov.}}{\Longrightarrow} \leq e^{-\lambda t} E\left[\exp\left(\lambda \frac{\sum x_i}{\|x\|_2}\right)\right]$$

$$\overset{\text{apply rademacher}}{\Longrightarrow} e^{-\lambda t} \underset{x,\varepsilon}{E}\left[\exp\left(\lambda \frac{\sum \varepsilon x_i}{\underbrace{\sqrt{\sum (x_i \varepsilon_i)^2}}_{= \|x\|_2}}\right)\right] = e^{-\lambda t} E_x E_\varepsilon\left[\lambda \frac{\sum \varepsilon_i x_i}{|x|_2}\right] \Big| X_{i-n}\Big]$$

$$\Rightarrow = e^{-\lambda t} E_x\left[\prod_{i=1}^{n} E_\varepsilon\left[\exp\left(\lambda \frac{\varepsilon_i x_i}{\|x\|_2}\right)\right]\right] = e^{-\lambda t} E_x\left[\prod_{i=1}^{n} \frac{1}{2}\exp\left(\frac{\lambda x_i}{\|x\|_2}\right) + \frac{1}{2}\exp\left(\frac{-\lambda x_i}{\|x\|_2}\right)\right]$$

$$\frac{1}{2}e^{\lambda} + \frac{1}{2}e^{-\lambda} \leq e^{\lambda^2/2}$$

$$\leq e^{-\lambda t} E_x\left[\prod_{i=1}^{n} \underbrace{\exp\left(\frac{\lambda^2 x_i^2}{2\|x\|_2^2}\right)}_{\exp\left(\frac{\lambda^2 \cancel{\sum x_i^2}}{2 \cancel{\sum x_i^2}}\right)}\right] = e^{-\lambda t} e^{\lambda^2/2} \leq \exp\left(-\lambda t + \lambda^2/2\right)$$

$\Longrightarrow$ optimize on $\lambda \to$ $-t + \lambda = 0 \Rightarrow \lambda = t$

$$\leq \exp\left(-\tfrac{1}{2}t^2\right)$$ ● Proof completed!

$$C = \tfrac{1}{2}$$

Q3) $V \in R^n$ $\quad X: x_1 \dots x_n$ $\quad \leftarrow$ Setup

$\forall f \in F, \exists v \in V$ s.t $\left(\frac{1}{n} \sum_{i=1}^{n} |f(x_t) - v_t|^t \right)^{1/p} \leq \alpha$

$N_p(F, \alpha, x^n)$

a) $1 \leq p \leq 4 \leq \infty$

we want to prove that $N_p(F, \alpha, x^n) \leq N_4(F, \alpha, x^n)$

$$\left(\frac{1}{n} \sum_{i=1}^{n} |f(x_t) - v_t|^p\right)^{1/p} = \left(\left(\frac{1}{2} \sum |f(x_t) - v_t|^p\right)^{4/p}\right)^{1/4}$$

from convexity of x^a (when $a > 1$ & $\frac{4}{p} > 1$) we have.

$$\leq \left(\frac{1}{n} \sum_{t=1}^{n} \left(|f(x_t) - v_t|^p\right)^{\frac{4}{p}}\right)^{1/4} = \left(\frac{1}{n} \sum_{t=1}^{n} |f(x_t) - v_t|^4\right)^{1/4}$$

$\Longrightarrow$ every α-cover for l_4 could be α-cover for l_p

so imagin $|V| = N_4(F, \alpha, x_1^n)$

$\Rightarrow \forall f \in F \ \exists v \in V \quad \left(\frac{1}{n} \sum |f(x_t) - v_t|^4\right)^{1/4} \leq \alpha$

$\left(\quad {}^{p}, {}^{1/p} \leq \quad \Longrightarrow \right.$

V is α cover for l_p norm too

$$\Rightarrow \quad N_p(F, \alpha, x_1^n) \leq |V| \leq N_q(F, \alpha, x_1^n) \quad \blacksquare$$

b)

$$R(F) = E_\sigma \left[\sup_{f \in F} \left\{ \frac{1}{n} \sum \sigma_i f(x_i) \right\} \right]$$

assume $|V| = N_1(F, \alpha, x)$

Now create function G such that it maps every function $f \in F$ into $[-1, 1]$

So now we have

$$E_\sigma \left[\sup_f \left\{ \frac{1}{n} \sum \sigma_i \left(f(x) - g(f) + g(f) \right) \right\} \right] = E_\sigma \left[\sup_f \left\{ \frac{1}{n} \sum \sigma_i (f(x) - g(f)) + \frac{1}{n} \sum_{\sigma_i G(f)} \right\} \right]$$

$$\leq \underbrace{E_\sigma \left[\sup_f \left\{ \frac{1}{n} \sum \sigma_i (f(x_i) - g(f)) \right\} \right]} + E_\sigma \left[\sup_f \left\{ \frac{1}{n} \sum \sigma_i G(f) \right\} \right]$$

$$\leq \underbrace{E_\sigma \left[\sup_f \left\{ \frac{1}{n} \sum |f(x_i) - g(f)| \right\} \right]}_{\leq \alpha \text{ 2 by definition}} + \underbrace{E_\sigma \left[\sup_f \left\{ \frac{1}{n} \sum \sigma_i G(f) \right\} \right]}_{Q}$$

Jenson.

$$Q = \frac{1}{n} \frac{1}{\lambda} E_\sigma \left[\log \sup \left\{ e^{\lambda \sigma^T G} \right\} \right] \leq \frac{1}{n\lambda} \log \left(\sum_G E\left[\prod_{i=1}^n e^{\lambda \sigma_i G_i} \right] \right)$$

$$\leq \frac{1}{n\lambda} \log \left(\sum_G \prod_{i=1}^n \left(e^{\frac{\lambda^2}{2} G_i^2} \right) \right) = \frac{1}{n\lambda} \log \left(\sum_G e^{\frac{\lambda^2}{2} \left(\sum G_i^2 \right)} \right)$$

$$= \frac{1}{n\lambda} \log\left(\sum_{G} e^{\frac{\lambda^2}{2}\|G\|^2} \right) \leq \frac{1}{n\lambda} \log\left(|G| \sup_{G \in G(f)} e^{\frac{\lambda^2}{2}\|G\|^2} \right)$$

apply log

$$= \frac{1}{n\lambda} \log|V| + \frac{\lambda}{2n} \sup \|G\|^2$$

optimize on λ $\Rightarrow$ $\frac{-\log|V|}{n\lambda^2} + \frac{1}{2n} \sup_{G} \{\|G\|^2\} = 0$

$$\Rightarrow \lambda = \frac{\sqrt{2\log|V|}}{\sup_{G}\{\|G\|\}} \longrightarrow \quad G \in [-1,1] \quad \|G\| \leq \sqrt{n} \Rightarrow$$

finall $\rightarrow$ for every $\alpha \geq 0$

$$\Rightarrow \leq \alpha + \frac{1}{n}\sqrt{2}\sqrt{\sqrt{n}\log|V|} = \alpha + \sqrt{\frac{2\log N_1(F,\alpha)}{n}}$$

for every $\alpha \Longrightarrow$

$$= \inf_{\alpha \geq 0} \left\{ \alpha + \sqrt{\frac{2\log N_1(F,\alpha)}{n}} \right. \qquad \blacksquare$$

---

c) we know that if $1/p + 1/q = 1$ then

$$\|x^T y\| \leq \|x\|_p \|y\|_q$$

$$\Rightarrow \begin{matrix} \|f\|_\infty = 1 \\ |x|_1 \end{matrix} \Rightarrow |\langle f, x\rangle| \leq \|f\|_\infty |x|_1 \leq |x|_1$$

Imagine $U \in \mathbb{R}^d$ be a α-cover L_∞ for F.

now create set W where $w = \sum_{i=1}^{n} \langle u, x_i \rangle \hat{e}_i$

↳ unit vector

$\Rightarrow G : f \to W$

$\forall f : |f - G(f)|_\infty = \max |f_i - G(f)| \leq \alpha$

now define V as $G(U)$ and $H : f \to V$

$$H(f) = \sum_i \langle G(f), x_i \rangle \hat{e}_i$$

now for ever $f \in F$ we would have.

$$\lim_{q \to \infty} \left(\frac{1}{n} \sum_i |f(x_i) - H(f)|^q \right)^{1/q} = \max_i |f(x_i) - G(f)$$

$$= \max_i \underbrace{|\langle f - H(f), x_i \rangle|}_{\leq \alpha} \leq \alpha$$

$\Rightarrow$ if we consider α-cover with L_∞ on V we found an α cover on F

$[-1, 1] \longrightarrow$ seperate by 2α each edge

$\Rightarrow$ consider nodes of all cubes as covering set.

$$|U| = \left(\frac{1-(-1)}{2\alpha}\right)^d = \left(\frac{1}{\alpha}\right)^d$$

and we know from lemma a that $N_1 < N_\infty \leq |U| \leq \left(\frac{1}{\alpha}\right)^d$

from lemma b $\rightarrow$ $\hat{R}[f] \leq \inf_{\alpha \geq 0} \left\{ \alpha + \sqrt{\frac{2 \log N_1}{n}} \right\}$

$$\inf_{\alpha \geq 0} \left\{ \alpha + \sqrt{\frac{2 \log (1/\alpha)^d}{n}} \right. = \alpha + \sqrt{\frac{2d \log 1/\alpha}{n}}$$

$$\alpha + \sqrt{\frac{2d \log 1/\alpha}{n}} \simeq O\left(\sqrt{\frac{d \log n}{n}}\right) \Rightarrow \alpha = 1/n \rightarrow$$

$$\inf_{\alpha \geq 0} (\quad) \leq \sqrt{\frac{2d \log n}{n}} + 1/n \approx O\left(\sqrt{\frac{d \log n}{n}}\right) \blacksquare$$

---

d) as we learned in class define $\alpha_i = 2^{-i}$ and $V_i \in \mathbb{R}^n$

an α_i-cover on $\{x_{-i}\}_{j=1}^{n}$ with l_2 norm.

because $y \in [-1,1]$ $\rightarrow$ for every $f \in F$ we have $|f(x_i)| \leq 1$

$\Longrightarrow$ if we consider $\vec{\Phi}$ then $\Rightarrow$

$$\forall f \in F : \left(\frac{1}{n} \sum_{i=1}^{n} \underbrace{|f(x_i) - V_i|^p}_{|f(x_i)|^p}\right)^{1/p} \leq 1$$

so V_0 only needs 1 point, we choos $r = \alpha_0 \leftrightarrow \alpha_L = 1 \leftarrow 2^{-L}$

$$\forall f \in F : \left(\frac{1}{n} \sum_{j=1}^{n} |f(x_j) - \underbrace{G_j(f)}_{\text{mapping}}|^2\right)^{\frac{1}{2}} \leq \alpha_i < 2^{-i}$$

$\Rightarrow \forall f: \; G_0(f) = 0$

chaining here.

$$\hat{R}(F) = E_\sigma\left[\sup_{f\in F}\left\{\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(x_i)\right\}\right]$$

$$\Rightarrow \hat{R}(F) = E_\sigma\left[\sup_f\left\{\frac{1}{n}\sum\sigma_i\left(f(x_i)-G^L(f)\right) + \frac{1}{n}\sum_{i=1}^{n}\sigma_i\sum_{j=1}^{L}G^j(f)-G^{j-1}(f)\right\}\right]$$

$$\leq E_\sigma\left[\sup_f\left\{\left(\frac{1}{n}\left(\sum f(x_i)-G^L(f)\right)^2\right)^{1/2} + \frac{1}{n}\sum_{j=1}^{L}\sigma^T\left(G^j(f)-G^{j-1}(f)\right)\right\}\right]$$

$$\leq E_\sigma\left[\sup_f\left\{\alpha_L + \frac{1}{n}\sum_{j=1}^{L}\sigma^T\left(G^j(f)-G^{j-1}(f)\right)\right\}\right]$$

we know that $\|G^j(f)-G^{j-1}(f)\|_2 \leq \alpha_{j-1}+\alpha_{j-2}$

$\leq 3\times 2^{-j}$

$$\hat{R}(F) \leq \alpha_L + \frac{1}{n}\sum_{j=1}^{L}E_\sigma\left[\sup\left\{\sigma^T\left(G^j(f)-G^{j-1}(f)\right)\right\}\right]$$

by-exp-Jensen

$$\leq \alpha_L + \frac{1}{n}\sum_{j=1}^{L}\frac{1}{\lambda_j}\log\left(\sum_{j\leq j-1}E_\sigma\left[\prod e^{\lambda_j\sigma_i\left(G^j(f)-G^{j-1}(f)\right)}\right]\right)$$

$$\leq \alpha_L + \frac{1}{n}\sum\frac{1}{\lambda_j}\log\left(\sum_{j,j-1}\exp\left(\frac{\lambda_j^2}{2}\sum\left(G^j(f)-G^{j-1}(f)\right)^2\right)\right)$$

$$\leq \alpha_L + \frac{1}{n}\sum_{j=1}^{L}\frac{\log(|G^j|)}{\lambda_j} + \frac{\lambda_j}{2}\sup_{G^j,G^{j-1}}\left\{\|G^j(f)-G^{j-1}(f)\|_2^2\right\}$$

$\Rightarrow$ Optimize on $\lambda_j \rightarrow$ $\lambda_j = \dfrac{\sqrt{2\log(|G^j|)}}{\sup_f\{\|G^j-G^{j-1}\|_2\}}$

$$\Rightarrow \leq \alpha_L + \frac{1}{n} \sum \sqrt{2^j \log |G^j|} \; \underbrace{\sup \|G^j - G^{j-1}\|_2}_{3 \times 2^{-j} \times \sqrt{n}}$$

$$\Rightarrow \alpha_L + \frac{6}{\sqrt{n}} \sum_i \sqrt{\log N_2(F, \alpha_j)} \; \alpha_j$$

$$\leq 2\alpha_L + \frac{12}{\sqrt{n}} \int_{\alpha_{L+1}}^{1} \sqrt{\log N_2(F, \delta)} \, d\delta$$

Now set L like $\alpha \leq \alpha_{L+1} = 2^{-L-1} \leq 2\alpha \Rightarrow 2\alpha_{L+1} \leq 4\alpha$

$$\int_{\alpha_{L+1}}^{1} \bigcirc \leq \int_{\alpha}^{1} \bigcirc$$

$$\Rightarrow \hat{R}\{F\} \leq 4\alpha + \frac{12}{\sqrt{n}} \int_{\alpha}^{1} \sqrt{\log N_2(F, \delta)} \, d\delta$$

$$\Rightarrow \hat{R}[F] \leq \inf_{\alpha \geq 0} \left\{ 4\alpha + \frac{12}{\sqrt{n}} \int_{\alpha}^{1} \sqrt{\log N_2(F, \delta)} \, d\delta \right\}$$

Proof completed.

---

e) $[-1,1] \to$ divide it to parts with length $\alpha \to$

$U = \{ [k_i \alpha], \; k \in \mathbb{Z} \wedge -\frac{1}{\alpha} < k \leq \frac{1}{\alpha} \} \to$ number of this points

$\lceil \frac{2}{\alpha} \rceil^n \approx (\frac{2}{\alpha})^n$

by this covering we have. $\left(\frac{1}{n}\sum_i |f(x_i) - G(f)_i|^p\right)^{\frac{1}{p}} \leq \alpha$

where we define. $G(f)_i = \left\lfloor \frac{f(x_i)}{\alpha} \right\rfloor \times \alpha$

now if we find number of $U \rightarrow$

on n point. $\rightarrow$ they are atleast 1 and max $\min\{n, \frac{2}{\alpha}\}$

now we have $\rightarrow$ from $U \rightarrow$ choose function mapping

then we would have. at most k distinct value.

$\Rightarrow$ x_1 . . . x_n

$\rightarrow \binom{n-1}{k-1} \rightarrow$ choose for their values.

$$\Rightarrow |U| \leq \sum_{k=1}^{n} \binom{\lfloor \frac{2}{\alpha} \rfloor}{k} \binom{n-1}{k-1} = \binom{\frac{2}{\alpha} + n - 1}{n}$$

$$\leq n^{2/\alpha}$$

$$\Rightarrow N(F_\alpha, x^n)_{\infty} \leq (n)^{\frac{2}{\alpha}}$$

for bound b) $\rightarrow$ $\inf_{\alpha \geq 0} \left\{ \alpha + \sqrt{\frac{2 \log(n^{2/\alpha})}{n}} \right\} = \left\{ \alpha + 2\sqrt{\frac{\log n}{\alpha n}} \right\}$

optimize $\rightarrow$ $1 + \cancel{2}\sqrt{\frac{\log n}{n}}\, \cancel{\frac{1}{2}}\, \alpha^{-3/2} = 0 \Rightarrow 1 = \sqrt{\frac{\log n}{n}}\, \alpha^{-3/2} \Rightarrow$

$\alpha^{3/2} = \sqrt{\frac{\log n}{n}} \Rightarrow \alpha = \left(\frac{\log n}{n}\right)^{1/3}$

$$\Rightarrow \alpha = \left(\frac{\log n}{n}\right)^{3} \rightsquigarrow \left(\frac{\log n}{n}\right)^{1/3} + 2\sqrt{\frac{\log n}{\left(\frac{\log n}{n}\right)^{1/3} n}} =$$

$$\left(\frac{\log n}{n}\right)^{1/3} + 2\left(\frac{\log n}{n}\right)^{1/3} = \boxed{3\sqrt[3]{\frac{\log n}{n}}} \qquad 3\left(\frac{\log n}{n}\right)^{1/3} \approx O\left(\left(\frac{\log n}{n}\right)^{1/3}\right)$$

for bound d_i:

$$\inf_{\alpha \geq 0}\left\{4\alpha + \frac{12}{\sqrt{n}}\int_{\alpha}^{1}\sqrt{\log N_2(F,\delta)}\,d\delta\right\} = \left\{4\alpha + \frac{12}{\sqrt{n}}\int_{\alpha}^{1}\sqrt{\log n^{2/\delta}}\,d\delta\right\}$$

$$\left\{4\alpha + \frac{12}{\sqrt{n}}\int_{\alpha}^{1}\sqrt{\frac{2}{\delta}}\sqrt{\log n}\,d\delta\right\} = \inf_{\alpha \geq 0}\left\{4\alpha + \frac{12\sqrt{2\log n}}{\sqrt{n}}\left(2 - 2\sqrt{\alpha}\right)\right\}$$

optimize on α

$$\cancel{4} - \cancel{12}^{\,3}\sqrt{\frac{2\log n}{n}} \times \frac{1}{\sqrt{\alpha}} = 0 \Rightarrow \sqrt{\alpha} = 3\sqrt{\frac{2\log n}{n}}$$

$$\alpha = \frac{18\log n}{n}$$

$$\text{bound} \leq \cancel{\frac{72\log n}{n}} + \frac{12\sqrt{2}}{\sqrt{n}}\sqrt{\log n}\left(2 - \cancel{6\sqrt{\frac{2\log n}{n}}}\right)$$

$$\text{bound} \leq 24\sqrt{\frac{2\log n}{n}} \qquad \approx O\left(\sqrt{\frac{\log n}{n}}\right)$$

this bound is tighter than $\left(\frac{\log n}{n}\right)^{1/3}$

f) fat shattering means that the biggest number of points in size that functions from class F can learn any labeling on them with safe margin at least $\frac{\alpha}{4}$

additionally it indicates $g(x_i) = \begin{cases} \geq r_i + \alpha & \text{if } y_i = 1 \\ \leq r_i - \alpha & \text{if } y_i = -1 \end{cases}$

---

g) we divide $[-b, b]$ into intervals with length $\frac{\alpha}{2}$

$\Rightarrow$ we will have $\frac{4}{\alpha}$ intervals.

now imagin $\frac{4}{\alpha}+1$ point there should be two point that are in the same interval (Pigen on hole principle)

$\Rightarrow$ any $\frac{4}{\alpha}+1$ point can not be shattere with safe margin α $\Rightarrow$ fat $< \frac{4}{\alpha}+1 \Rightarrow$ $fat(F, \alpha) \leq \frac{\alpha}{4}$ •

---

h) we know that $N_2(F, \alpha) \leq N_\infty(F, \alpha)$

assume α-cover on x_i^n with norm L_∞ (U) this cover is

$\forall f$ ther is u in U : $\lim_{p \to \infty} \left(\frac{1}{n} \sum_i |f(x_i) - u_i|^p \right)^{1/p} = \max_{i=[n]} \|f(x_i) - u_i\| \leq \alpha$

now for each point x_i we have.

$x: \rightarrow$ we find $f, g \in F$ that $\sup\{|f(x_i) - g(x_i)|\} \leq \alpha$

which means that $|f(x_i) - u_i| \leq \alpha \leftarrow \forall f \in F \ \exists u \in U$

or $\exists f, g$ that $|f(x_i) - g(x_i)| > \alpha \rightarrow$ consider $w = \left[\frac{f(x_i)}{\alpha}\right]\alpha$

$\Rightarrow |w_i - f(x_i)| \leq \alpha \Rightarrow$ we need $\left(\frac{2}{\alpha}\right)$ for w_i

$\Rightarrow$ we have $fat(F, \alpha)$ most number of point situation

happens $\Rightarrow N_\infty(F, \alpha) \leq \left(\frac{2}{\alpha}\right)^{fat(F)} \Rightarrow$

$$N_2(F, \alpha) \leq \left(\frac{2}{\alpha}\right)^{fat(F)}$$

we can have k value for $f \in F \Rightarrow \left(\frac{2}{\alpha}\right)^{k\, fat(F, c\alpha)}$

as long as $ck = 1$

---

from before

1) $$\hat{R}(F) \leq \inf_{\alpha}\left\{4\alpha + \frac{12}{\sqrt{n}}\int_{\alpha}^{1}\sqrt{\log \underbrace{N_2(F, \delta)}_{\left(\frac{2}{\delta}\right)^{k\, fat(F, \alpha c)}}}\, d\delta\right\}$$

$$\hat{R}(F) \leq \inf_{\alpha \geq 0}\left\{4\alpha + \frac{12}{\sqrt{n}}\int_{\alpha}^{1}\sqrt{k\, fat(F, c\delta)\log\frac{2}{\delta}}\, d\delta\right\}$$

Q4)

a) imagine set U $\{u_1 \dots u_k\}$ that can be shattered by $\bigcup_{i=1}^{n} H_i$ so this implies that $|\{h \cap \{x_1 \dots x_k\} : h \in \bigcup H_i\}|$

$$= 2^{vcdim(\bigcup_{i=1}^{n} H_i)} = 2 \rightarrow \text{number of labeling.}$$

$$\Rightarrow |\{h \cap \{x_1 \dots x_k\} : h \in \bigcup H_i\}| = \bigcup_{i=1}^{n} \{h \cap \{x_1 \dots\} \mid h \in H_i\}$$

$$2^{vc} = \left|\bigcup_{i=1}^{n} \{h \cap \{x \dots x_k\} \mid h \in H_i\}\right| \overset{\text{union bound}}{\leq} \sum_{i=1}^{n} |\{h \cap \{x_1 \dots x_k\} \mid h \in H_i\}|$$

$$\leq \sum_{i=1}^{n} \sum_{j=0}^{vc(h_i)} \binom{vc}{j} \leq \sum_{i=1}^{n} \sum_{j=0}^{d \to max} \binom{k}{j} \leq n \times \sum_{j=1}^{d} \binom{vc}{j}$$

$d \leq vc$, $d \geq e$

$$\leq n \left(\frac{e\, vc}{d}\right)^d \leq n\, vc^d \Rightarrow$$

$$2^{vc} \leq n\, vc^d \Rightarrow vc \leq \lg n + d \lg vc$$

$$r = n \Rightarrow vc \leq \lg r + d \lg vc$$

$$f(n) = n - d \lg n$$

so $\rightarrow$ $f'(x) = 1 - \frac{d}{n} \Rightarrow n = d$

$$1 - \frac{d}{3d \lg d} = \frac{2d \lg d - d}{3d \lg d} \geq 0$$

$$f'(2d \lg d) \geq 0$$

$$f(2d \lg d) \rightarrow 2d \lg d - d \lg 3d \lg d \geq 0$$

$$\Rightarrow f(2d \lg d) \geq 0 \quad , \quad f'(2d \lg d) \geq 0$$

$\Rightarrow \forall d > 0 \mid x \leq d \lg vc \Rightarrow vc < 2d \lg d$

$\Longrightarrow \quad vc \leq 4 \lg 2d + 2 \lg r$

set $d = 2d$

Now imagine $vc \geq 2d \lg d \rightsquigarrow vc \geq 4d \lg 2d$

$\Rightarrow \quad vc \leq 2 \lg r + 4d \lg 2d$ ↙ from defro ∎

Part b next page.

b) if $r=2 \rightsquigarrow$ no set with size $2d+2$ can be shattered.

$$|h \cap \{x_1 \dots x_{2d}\} \mid h \in H_1 \cup H_2| = \sum_{i=1}^{2} |h \cap \{x_1 \dots x_{2d}\} \mid h \in H_i\}$$

$$\leq \sum_{j=0}^{vc(H_1)} \binom{2d+2}{j} + \sum_{i=0}^{vc(H_2)} \binom{2d+2}{i} \leq 2^{2d+2} - O(2^d)$$

$$\Rightarrow 2^{vc} < 2^{2d+2} \Rightarrow vc < 2d+2$$

Q5) a) if $k=2$ then if set U shattered by N dim then VC hattered too so if set U is shattered means

$\forall x \in U : f_0(x) \neq f_1(x)$

$\exists f_0, f_1$

$\forall V \subseteq U \ \exists h \in H : \forall x \in V \ h(x) = f_0(x) \wedge \forall x \notin V \ h(x) = f_1(x)$

we want to proof that this would be shattered in terms of VC $\rightarrow$ imagine any labelling on x_1^n such y_1^n

$\Rightarrow [0,1] \Rightarrow f_0(x) = 1 - f(x) \Rightarrow$ if $y_1 = f_0(x)$ or $y_i = f_1(x)$ we find the function. or $V \subseteq U \rightarrow y_i = f(x_i)$ and $v \notin V$ $y_1 = f_1(x_i) \Rightarrow$ implies there exist a function $h \in H$ which is lableing function. $\Rightarrow$ VC dim $\geq$ N dim

1

because every set shattered in terms of Ndim is shattered in terms of VC.

$k=2 \rightarrow$ then we have $VC \leq Ndim \rightarrow$ if any set shatter by VC shatters by Ndim. $\rightarrow$

for every labelling $\{y_i\} \rightarrow U = \{x_i \in U : y_i = 0\} \rightarrow U \setminus V = \{x_i \in U : y_i = 1\}$

$f_0(x) = 0$ and $f(x) = 1 \rightarrow$ for every $x_i \Rightarrow f_0(x_i) \neq f_1(x_i)$

$\Rightarrow \forall x \in V \quad h(x) = f_0(x)$, $\forall x \in U/V \quad h(x) = 1 = f_1(x)$

$\Rightarrow$ there is a function $h \in H$ that satisfies Ndim conditions. $\Rightarrow$ $VC \leq Ndim$ (2)

(1), (2) $\Rightarrow$ $VC = Ndim$

---

b) we use induction on $|X| \rightarrow$ base $|X| = 1$

$\Rightarrow$ for $|H| = 1 \Rightarrow Ndim(H) \geq 0$ and

$$|X|^{Ndim(H)} \quad K^{2Ndim(H)} \geq 1 = |H|$$

if $|H| > 1 \rightarrow$ we show $Ndim(H) \geq 1$

$H =$ atleast $h_1, h_2 \rightarrow$ so there exist f_0, f_1 we have

$|X| = 1 \Rightarrow f_0(x) \neq f_1(x) \rightarrow \{X\} = \{\emptyset, \{x\}\}$

so X shatter by H.

$|H| \lesssim K \lesssim K^2$

$|X|^{N\dim(|H|)} K^{2N\dim(|H|)} \geq K^2$ $\Big\} \Rightarrow \|H\|^{N\dim(|H|)} K^{2N\dim(|H|)}$

induction hypothesis

induction step?

$|X| = m \to |X| = m+1$

$H_i = \{h \in H,\ h(x_1) = i\}$

$H_{ij} := \{h \in H_{ij} \ \exists h' \in H_j \mid \forall x \quad h'(x) = h(x)\} \quad i \neq j$

$\hat{H} = H - \bigcup_{i \neq j} H_{ij}$

$\Rightarrow |H| = |\hat{H}| + |\bigcup_{i \neq j} H_{ij}| \leq |\hat{H}| + \sum_{i \neq j} |H_{ij}|$

$N^{\dim(|\hat{H}|)} + K^{2N\dim(|\hat{H}|)}$

$|H_{ij}| \overset{?}{\lesssim} N^{\dim(H_{ij})} K^{2N\dim(H_{ij})} \longrightarrow$ Strong induction.

$\Rightarrow |H| \leq \underbrace{|X|^{\dim(|\hat{H}|)} + K^{2N\dim(\hat{H})}}_{C_1} + \sum_{i \neq j} |X|^{N\dim(H_{ij})} + K^{2N\dim(H_{ij})}$

$\dim(H_{ij}) \leq \dim(H)-1 \rightarrow$ trivial

$$|H| \leq C_1 + \sum \underbrace{|X|^{N\dim(H)-1} + K^{2N\dim(H)-2}}_{C_2}$$

$$|H| \leq C_1 + \binom{K}{2} C_2$$

$$|H| \leq C_1 + K^2 C_2 \leq |H|^{N\dim(H)} K^{2N\dim(H)} \quad \boxtimes$$

proof completed $\square$

a)

Q6) $P[\lambda_{max}(S_n) > \delta] = P[\varphi(\alpha \lambda_{max}(S_n)) \geq \varphi(\alpha\delta)]$

increasing. $\leftarrow \varphi$

$\leq E[\varphi(\lambda_{max}(\alpha S_n))]\, \varphi^{-1}(\alpha\delta)$

$S_n = \sum \lambda_i u_i u_i^T \rightarrow \varphi(\alpha S_n) = \sum \varphi(\alpha\lambda_i) u_i u_i^T$

$\Rightarrow \varphi(\lambda_{max}(\alpha S_n)) \leq \lambda_{max}(\varphi(\alpha S_n))$

$\lambda_{max}(\varphi(\alpha S_n)) = \varphi(\alpha \lambda^{max}) \leq \sum \varphi(\alpha\lambda_i) = Tr(\varphi(\alpha S_n))$

$$P[\lambda_{max}(S_n) > \delta] \leq \frac{E[Tr(\varphi(\alpha S_n))]}{\varphi(\alpha\delta)} = \frac{Tr(E[\varphi(\alpha S_n)])}{\varphi(\alpha\delta)}$$

$$P[\lambda_{max}(S_n) > \delta] \leq \frac{Tr(E[\varphi(\delta S_n)])}{\varphi(\alpha\delta)} \quad \blacksquare$$

---

b) $E[\exp(\alpha A)] = I + \sum_{n=2}^{\infty} \frac{\alpha^n E[A^n]}{n!}$

$A = \sum \lambda_i u_i u_i^T \Rightarrow A^n = \sum \lambda_i^n u_i u_i^T$ because $\|A\|_{op} \leq 1$ then

$\forall \lambda_i\ |\lambda_i| \leq 1 \Rightarrow \forall n \geq 2 \rightarrow \lambda_i^2 \leq |\lambda_i|^n \leq |\lambda_i|^2 = \lambda_i^2$

$\Rightarrow A^n - A^2 = \sum \underbrace{(\lambda^n - \lambda^2)}_{\leq 0} u_i u_i^T \Rightarrow A^n - A^2 \leq 0$

$\Rightarrow A^n \leq A^2$

$$E[\exp(\alpha A)] = I + \sum_{i=2}^{\infty} \frac{\alpha^n E[A^n]}{i!} \leq I + \sum_{i=2}^{\infty} \frac{\alpha^n E[A^2]}{i!}$$

$$= I + \mathrm{var}(A)\,(e^{\alpha} - \alpha - 1) = I + \varphi(\alpha)\,\mathrm{var}(A)$$

from defor $\quad 1+x \leq e^x$

$$\Rightarrow \quad \leq \quad e^{\varphi(\alpha)\,\mathrm{var}(A)} \quad \Rightarrow$$

$$E[\exp(\alpha A)] \leq \exp(\varphi(\alpha)\,\mathrm{var}(A)) \overset{\log}{\Longrightarrow}$$

$$\log E[\alpha A] \leq \log \exp(\varphi(\alpha)\,\mathrm{var}(A)) \Rightarrow \varphi(\alpha)\,\mathrm{var}(A)$$

---

c) $$P[\lambda_{max}(S_n) > \delta] \leq \frac{\mathrm{Tr}(E(\varphi(\alpha S_n)))}{\varphi(\alpha\delta)}$$

$$\frac{\mathrm{Tr}(E[e^{\alpha S_n} - \cancel{\alpha S_n} - I])}{\varphi(\alpha\delta)} = \frac{\mathrm{Tr}(E[e^{\alpha S_n} - I])}{\varphi(\alpha\delta)}$$

$$E[\mathrm{Tr}(\exp(H + \log X))] \leq \mathrm{Tr}(\exp(H + \log(E(X))))$$

$$\Rightarrow \mathrm{Tr}(E[e^{\alpha S_n}]) = E_{A_2 \ldots n}\, E_{A_1 | A_2 \ldots A_n}\, \mathrm{Tr}(\exp(\alpha \sum A_i + \log e^{\alpha A_i}))$$

$$\leq E_{A_1 \ldots A_n}[\mathrm{tr}(\exp(\alpha \sum A_i + \log E[e^{\alpha A_1}]))]$$

following thin.

$$\leq \operatorname{Tr}\left\{\exp\left(\sum_i \underbrace{\log E[e^{\alpha A_i}]}_{\leq \varphi(\alpha)\operatorname{var}(A_i)}\right)\right\}$$

$\Rightarrow$

$$\operatorname{Tr}(E[e^{\alpha S_n}]) \leq \operatorname{Tr}\left(\exp\left(\sum \varphi(\alpha)\operatorname{var}(A_i)\right)\right)$$

$$= \operatorname{Tr}(\exp(\varphi(\alpha)V))$$

$$\Rightarrow P(\lambda_{\max}(S_n) > \delta) \leq \frac{\operatorname{Tr}\{E[e^{\alpha S_n}] - I\}}{\varphi(\alpha\delta)}$$

$$\leq \frac{\operatorname{Tr}(\exp(\varphi(\alpha)V) - I\}}{\varphi(\alpha\delta)}$$

define $g(x) = e^x - 1 \rightarrow$ $\lambda \in \lambda(V)$

$$\operatorname{Tr}(\exp(\varphi(\alpha)V) - I) = \sum (e^{\varphi(\alpha)\lambda_i} - 1) = \sum_i (g(\varphi(\alpha)\lambda_i))$$

set $\beta_i = \frac{\lambda_i}{\|V\|_{op}} \Rightarrow 0 \leq \beta_i \leq 1$

g is convex

$$g(\varphi(\alpha)\lambda_i) = g(\varphi(\alpha)\beta_i\|V\|_{op}) \leq \beta_i\, g(\varphi(\alpha)\|V\|_{op})$$

$$\frac{\beta_i}{\|V\|_{op}}\left(e^{\varphi(\alpha)\|V\|_{op}} - 1\right) \leq \frac{\lambda_i}{\|V\|_{op}} \exp(\varphi(\alpha)\|V\|_{op})$$

$$\mathrm{Tr}(\exp(\varphi(\alpha) V - I)) = \sum_i g(\varphi(\alpha)\lambda_i)$$

$$\leq \sum_i \frac{\lambda_i}{\|V\|_{op}} \exp(\varphi(\alpha)\|V\|_{op}) = \frac{\mathrm{Tr}(V)}{\|V\|_{op}} \exp(\varphi(\alpha)\|V\|_{op})$$

$$\Rightarrow P\{\lambda_{max}(S_n) > \delta\} \leq \frac{\mathrm{Tr}(V)}{\|V\|_{op}} \inf_{\alpha \geq 0} \left\{ \frac{\exp(\varphi(\alpha)\|V\|_{op})}{\varphi(\delta\alpha)} \right\}$$

Proof completed. ✓

---

d) same as class define $Q = \begin{bmatrix} 0 & x^T \\ x & 0 \end{bmatrix}$

$E(Q) = 0$

from class we know that if x is rank one then Q's rank ≤ 2 then also we know $\lambda(Q) = \pm|x|_2$

so $\Rightarrow |Q|_{op} = \|x\|_2 \leq 1$

$$S_n = \begin{bmatrix} 0 & \sum X_i \\ \sum X_i & 0 \end{bmatrix} \Longrightarrow |\lambda(S_n)| = \|\sum_{k=1}^n X_i\|_2 \Rightarrow \|S_n\|_{op} = \lambda_{max}(S_n)$$

$$= \|\sum_{k=1}^n X_k\|_2 \Longrightarrow P\{\|\sum X_k\|_2 \geq \delta\} \leq \frac{\mathrm{Tr}(V)}{\|V\|_{op}} \inf_{\alpha \geq 0} \left\{ \frac{\varphi(\alpha)\|V\|_{op}}{\varphi(\alpha\delta)} \right\}$$

where $V = \sum E[Q_i^2]$

we have $Q^2 = \begin{bmatrix} \|X\|_2^2 & 0 \\ 0 & XX^T \end{bmatrix} \Rightarrow E[Q^2] =$

$$\begin{bmatrix} E[\|X\|_2^2] & 0 \\ 0 & E[XX^T] \end{bmatrix} = \begin{bmatrix} Tr(C) & 0 \\ 0 & C \end{bmatrix}$$

$$V = \sum E\{Q_i^2\} = n \begin{bmatrix} Tr(C) & 0 \\ 0 & C \end{bmatrix},$$

$$Tr(V) = 2n\, Tr(C)$$

$\|V\|_{op} = n\, Tr(C) \rightarrow$ from class.

$$P[\bigcirc] = \frac{2n\, Tr(C)}{n\, Tr(C)} \inf_{\alpha \geq 0} \left\{ \frac{\exp(\varphi(\alpha)\, n\, Tr(C))}{\varphi(\delta\alpha)} \right\}$$

$$= 2 \inf_{\alpha \geq 0} \left\{ \frac{\exp(\varphi(\alpha)\, n\, Tr(C))}{\varphi(\alpha\, \delta)} \right\}$$

Compare to Bernstein $\Rightarrow$

$$\bigcirc \leq 2\,\mathrm{rank}(V) \exp\left(\frac{-\delta^2}{2\|V\|_{op} + b\delta} \right)$$

the rank(V) can grow to $d+1$

but here the constant is $\underline{\underline{2}}$ $\boxtimes$

Q8) Van handel 7.1 :

a)

we know if $B = \{ \|X\|_2 \leq 1 \}$ then $g(B) = \sqrt{\lg n}$

$V(B) = 1$

$$\Rightarrow V(B) = \mathbb{E}_{\varepsilon}\left[\sup_{v \in B} \langle \varepsilon, v \rangle\right] \leq \mathbb{E} \sup \|X\| = \mathbb{E}[1] = 1 \blacksquare$$

$$g(B) = \mathbb{E}_{g}\left[\sup \sum g_i v_i\right] \leq \mathbb{E}_{g}\left[\max |g_i|\right] \sim \sqrt{\lg n} \blacksquare$$

$\hookrightarrow v_i = 1, a$

---

b)

$$T \subseteq T_1 + T_2 : \left\{ \begin{array}{l} g(T_1) \leq a \\ \sup_{t \in T_2} |t|_1 \leq a \end{array} \right\} \Rightarrow V(T) = \mathbb{E} \overbrace{\sup \langle \varepsilon, v \rangle}^{\leq a} \leq a \blacksquare$$

$$\mathbb{E}_{\varepsilon}\left[\sup_{v \in T} \langle \varepsilon, v \rangle\right] = \mathbb{E} \sup_{\substack{v_1 \in T_1 \\ v_2 \in T_2}} \left[\langle \varepsilon, v_1 \rangle + \langle \varepsilon, v_2 \rangle\right]$$

$$\leq \underbrace{\mathbb{E} \sup_{T_1} \langle \varepsilon, v_1 \rangle}_{V(T_1)} + \underbrace{\mathbb{E} \sup_{T_2} \langle \varepsilon, v_2 \rangle}_{\sup |v|_1} = a + a \leq 2a \blacksquare$$

---

c) $|a_1| \dots |a_n| \leq 1 \Longrightarrow$

$$E \sup \sum \varepsilon_k v_k a_k \rightsquigarrow \overset{a_i = \pm 1 \to \text{edge} \Rightarrow}{}$$

$$E \sup \sum a_k v_k \varepsilon_k \leq E \sup \left[\sum v_k \varepsilon_k \right] \quad \blacksquare$$

---

d) $a_i = \dfrac{g_i}{\varepsilon_i} \longrightarrow \max a_i \leq \sqrt{\log n}$

$$\longrightarrow E\left[\sup g_i v_i\right] \leq \sqrt{\log n}\, E \sup \sum \varepsilon_i v_i$$

$$\Longrightarrow \quad g(T) \leq \sqrt{\log n}\, \gamma(T) \quad \blacksquare$$