تمرین تئوری ۲ یادگیری ماشین

نویسنده:

سید احسان حسن بیگی - ۴۰۲۲۱۱۷۲۳

پرسش ۱

$\boxed{1}$

به صورت کلی نشان می دهیم که $d_{vc}(H) = d+1$

perceptron

lower bound ابتدا نشان می دهیم مثالی وجود دارد که $d+1$ سمپل را در فضای $\mathbb{R}^{d+1}$ بچینیم، H بتواند آن را shatter کند

یعنی به ازای تمام 2^{d+1} حالت ممکن برای لیبل ها، H بتواند سمپل ها را دسته بندی کند

مجموعه سمپل ها (X) را به صورت زیر در نظر می گیریم

$$X = \begin{bmatrix} — x_1^T — \\ — x_2^T — \\ \vdots \\ — x_{d+1}^T — \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix}_{(d+1)\times(d+1)}$$

برای بایاس یک 1 به ابتدای هر سمپل d بعدی اضافه کردیم

حال به ازای هر 2^{d+1} حالتی که برای $y = \begin{bmatrix} \pm 1 \\ \vdots \\ \pm 1 \end{bmatrix}$ متصور باشیم وزن (w) ای وجود خواهد داشت به طوریکه $w = X^{-1}y$

زیرا مثال X را به صورتی انتخاب کردیم که X وارون پذیر باشد $\longrightarrow$ هر 2^{d+1} حالت دسته بندی شد

$$d_{vc}(H) \geq d+1 \quad (\mathrm{I})$$

upper bound حال باید نشان دهیم که اگر d+2 سمپل را در فضای $\mathbb{R}^{d+1}$ بچینیم، به هر نحوی که این چیدمان را انجام دهیم، H نمی‌تواند تمام 2^{d+2} حالت ممکن برای لیبل ها را دسته بندی کند.

از جبر خطی می‌دانیم که اگر d+2 نقطه را در فضای d+1 بعدی بچینیم حداقل یکی از نقاط، linear combination سایر نقاط خواهد بود و با توجه به اینکه المان اول تمام x_i ها را 1 در نظر گرفتیم، x_i ها و بنابراین ضرایب، نمی‌توانند تمام صفر باشند

$$x_j = \sum_{i \neq j} \alpha_i x_i \longrightarrow w^T x_j = \sum_{i \neq j} \alpha_i w^T x_i$$

$$\vec{\alpha} \neq \vec{0}$$

حال حالتی را در نظر می‌گیریم که لیبل x_i ها $\text{sign}(\alpha_i)$ باشد، در این حالت $w^T x_i$ و α_i هم علامت می‌شوند و بنابراین $w^T x_j$ حتماً مثبت می‌شود و بنابراین لیبل x_j نمی‌تواند منفی شود ← نشان دادیم که H نمی‌تواند d+2 سمپل را shatter کند

$$d_{VC}(H) < d+2 \quad \text{II}$$

$$\text{I}\ \text{II} \Longrightarrow d_{VC}(H) = d+1 \xrightarrow{d=3}$$ پرسپترون در $\mathbb{R}^3$ $d_{VC} = 4$ $\longrightarrow$ $\boxed{\text{کوچکترین نقطه شکست پرسپترون در } \mathbb{R}^3 = 5}$

⊛ فرض کردیم که منظور از فضای $\mathbb{R}^3$ آن است که فیچرها بدون در نظر گرفتن 1 ای که برای بایاس اضافه می‌کنیم، 3 بعد دارند

$\boxed{2.a}$

می خواهیم نشان دهیم که $d_{VC}(H) = N$

parity functions

lower bound

ابتدا نشان می دهیم مثالی وجود دارد که n سمپل را در feature space بچینیم و H بتواند آن را shatter کند

یعنی به ازای تمام 2^N حالت ممکن برای لیبل ها، H بتواند سمپل ها را دسته بندی کند

سمپل ها را $e_1 \dots e_n$ در نظر می گیریم به صورتی که $e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$ (i امین المان) . حال به ازای هر 2^N حالت ممکن برای لیبل ها اگر $e_{k_1} \cdots e_{k_j}$

لیبل مثبت گرفته باشند $h_S ; S=\{k_1 \dots k_j\}$ فقط المان های $k_1 \cdots k_j$ سمپل ها را XOR می کند و بنابراین فقط سمپل های $e_{k_1} \cdots e_{k_j}$

هستند که آن المان 1 شان شمرده می شود ← $e_{k_1} \cdots e_{k_j}$ ← لیبل مثبت

باقی سمپل ها ← لیبل منفی

هر 2^N حالت دسته بندی شد

$d_{VC}(H) \geq N$ (I)

upper bound

طبق تعریف صورت سوال از parity function می دانیم که $|H|$ بی نهایت نبوده و برابر تعداد S های ممکن است

با توجه به اینکه $S \subseteq \{1 \dots N\}$ پس $|H| = 2^N$

حال اگر $N+1$ سمپل انتخاب کنیم، تعداد تمام حالات ممکن برای لیبل ها 2^{N+1} می شود که از $|H|$ بزرگتر است

بنابراین امکان ندارد که H بتواند $N+1$ سمپل را shatter کند (به هر نحوی که چیده شده باشند)

$d_{VC}(H) < N+1$ (II)

$$\text{(I) (II)} \Longrightarrow \boxed{d_{VC}(H) = N}$$

$\boxed{2.b}$

می خواهیم نشان دهیم که $d_{VC}(H) = 2d$

axis aligned rectangles in $\mathbb{R}^d$

lower bound

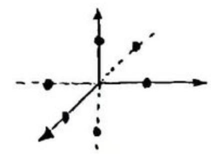
ابتدا نشان می دهیم مثالی وجود دارد که $2d$ سمپل را در $\mathbb{R}^d$ بچینیم و H بتواند آن را shatter کند

یعنی به ازای تمام 2^{2d} حالت ممکن برای لیبل ها، H بتواند سمپل ها را دسته بندی کند.

سمپل ها را طوری در نظر می گیریم که هرکدام آنها یکی از d محور مختصات را انتخاب کرده و در موقعیت $+1$ یا -1 آن قرار گیرد

$\{\pm e_1 \dots \pm e_d\}$

به عنوان مثال برای $d=3$ داریم



باتوجه به اینکه هرکدام از سمپل ها به این صورت بر روی محور ها قرار گرفته اند واضح است که هر زیر مجموعه‌ای از آن ها را می توانیم با یک مستطیل عضو H بگیریم

$$d_{VC}(H) \geq 2d \quad \text{(I)}$$

upper bound

در تعیین حد پایین سعی کردیم نقاط را به گونه ای پخش کنیم که هیچ نقطه‌ای بین نقاط دیگر قرار نگیرد زیرا در این صورت مستطیلی که نقاط بیرونی را در بر بگیرد مجبور است که نقطه بین آنها را نیز در بر بگیرد و بنابراین نمی توان تمام 2^{2d} حالت را داشت

حال باتوجه به اینکه $2d$ محور مختصات داریم اما $2d+1$ نقطه داریم، به هر صورتی که نقاط را بچینیم، یک نقطه بین نقاط دیگر می افتد

بنابراین طبق دلیل ذکر شده اگر قرار باشد نقطه داخلی لیبل منفی گرفته و سایر نقاط لیبل مثبت بگیرند، این حالت دسته بندی نخواهد شد

$$d_{VC}(H) < 2d+1 \quad \text{(II)}$$

$$\text{(I)}\,\text{(II)} \Longrightarrow \boxed{d_{VC}(H) = 2d}$$

$\boxed{2.c}$

طبق صورت سوال F یک function space است با خروجی حقیقی برای توابع

و $\dim(F) = d$ یعنی درجه توابع عضو F برابر d است

بنابراین می دانیم که F از نظر خطی بسته است چون باید خواص vector space را دارا باشد. پس:

$$\left.\begin{array}{l}\forall f_1, f_2 \in F \\ \forall r \in \mathbb{R} \text{ (فیلد)} \\ \forall x \in \mathbb{R}^n\end{array}\right\} \quad (f_1 + r f_2)(x) = f_1(x) + r f_2(x)$$

چون $\dim(F)$ را d گرفتیم، ورودی توابع یا همان تعداد سمپل‌ها را n در نظر می‌گیریم

ابتدا ثابت می کنیم که VC dimension برای $\text{sign}(f+g)$ برابر VC dimension برای $\text{sign}(f)$ است

برای این کار ابتدا ثابت می کنیم، اگر $\text{sign}(f)$ بتواند m تا سمپل را shatter کند حتماً $\text{sign}(f+g)$ نیز می تواند این کار را بکند (به ازای هر g حقیقی)

واضح است که f می تواند این کار را بکند. به عنوان مثال می توان f را ترکیب خطی $u, v \in F$ دانست به صورتی که

$$\begin{cases} u(x_i) > 0 \longleftrightarrow y_i = 1 \\ v(x_i) > 0 \longleftrightarrow y_i = -1 \end{cases}$$

حال چون F خواص vector space را دارد می توان گفت $f = u - v$

برای اینکه ثابت کنیم $\text{sign}(f+g)$ نیز می تواند m سمپل را shatter کند می توانیم به ازای هر g، f را طوری با constant مثبت amplify کنیم که به ازای هر سمپل $|f(x_i)| > |g(x_i)|$ و بنابراین صرف نظر از آن که g چی باشد علامت f عوض نخواهد شد

$$\text{sign}\big((f+g)(x_i)\big) = \text{sign}(f(x_i)) = y_i$$

برای جهت برعکس ثابت می کنیم: اگر $\text{sign}(f+g)$ بتواند m سمپل را shatter کند حتماً $\text{sign}(f)$ نیز می تواند

$u \in F$ ای انتخاب می کنیم که $\text{sign}\big((u+g)(x_i)\big) = y_i$ یعنی به درستی تمام سمپل‌ها را shatter کند

$v \in F$ ای انتخاب می کنیم که $\text{sign}\big((v+g)(x_i)\big) = -y_i$ یعنی برای هر سمپل لیبل اش را برعکس قرار دهد

حال می دانیم که $\text{sign}\Big(\big((u+g) - (v+g)\big)(x_i)\Big) = y_i$

دلیل آن این است که $v+g$ برعکس لیبل می زند پس $-(v+g)$ درست لیبل می زند ضمناً اگر 2 تابع که درست لیبل می زنند را جمع کنیم همچنان sign یکسانی خواهیم داشت

$\longleftarrow$ $u+g-(v+g) = u - v \in F$ پس نشان دادیم که اعضای F (به صورت مختصر f) نیز می توانند m سمپل را shatter کنند

بنابراین به طور کامل اثبات شد که $d_{VC}(\text{sign}(f)) = d_{VC}(\text{sign}(f+g))$

حال ثابت می‌کنیم d سمپل وجود دارد که shatter شوند اگر و تنها اگر زیرمجموعه d عضوی از F وجود داشته باشد که اعضای آن یعنی $f_1 \dots f_d$ مستقل خطی باشند

اگر این اثبات را انجام دهیم خواهیم داشت $d_{VC}(sign(F)) = dim(F) = d$

فرض می‌کنیم $f_1 \dots f_d$ مستقل خطی باشند و به ازای هر سمپل $x \in \mathbb{R}^n$ تعریف می‌کنیم $\phi(x) = (f_1(x), \dots, f_d(x))$ یعنی ϕ هر سمپل n بعدی را به d بعد می‌برد. حال با استقرا می‌توان نشان داد d سمپل $x_1 \dots x_d$ وجود دارد به صورتی که $\phi(x_1) \dots \phi(x_d)$ در فضای $\mathbb{R}^d$ مستقل خطی باشند و تمام فضا را span کنند.

برای حالت پایه داریم که $d=1$ و واضح است که یک سمپل مستقل خواهد بود. حال فرض می‌کنیم $\phi(x_1) \dots \phi(x_{d-1})$ مستقل خطی اند.
اگر x_d جدیدی وجود نداشته باشد که $\phi(x_1) \dots \phi(x_d)$ مستقل خطی باشند باید f_d از ترکیب خطی $f_1 \dots f_{d-1}$ ایجاد شود
با توجه به اینکه فرض در این سمت از اثبات آن بود که $f_1 \dots f_d$ مستقل خطی اند، پس فرض خلف باطل است و x_d نیز وجود خواهد داشت.

حال با توجه به اینکه F از نظر خطی بسته است می‌توان نوشت:

$$\left\{\left(sign(\langle w,\phi(x_1)\rangle), \dots, sign(\langle w,\phi(x_d)\rangle)\right) ; w \in \mathbb{R}^d\right\} \subseteq sign(F)_{\{x_1 \dots x_d\}}$$

d عضو مستقل در فضای d بعدی حکم basis را خواهند داشت و بنابراین می‌توان این basis ها یعنی $\phi(x_d) \dots \phi(x_1)$ را معادل basis های استاندارد یعنی $e_1 \dots e_d$ در نظر گرفت و می‌دانیم که پایه‌های استاندارد را می‌توان shatter کرد. پس نشان دادیم که d سمپل وجود دارد که shatter شوند ⟵ یک سمت را اثبات کردیم

برای سمت دیگر فرض می‌کنیم زیرمجموعه d عضوی از F نداریم و به جای آن $f_1 \dots f_k$ را پایه در نظر می‌گیریم $(k < d)$ و تعریف می‌کنیم $\phi(x) = (f_1(x), \dots, f_k(x))$

حال اگر d سمپل $x_1 \dots x_d$ را در نظر بگیریم داریم:

$$sign(F)_{\{x_1 \dots x_d\}} \subseteq \left\{\left(sign(\langle w,\phi(x_1)\rangle) \dots sign(\langle w,\phi(x_d)\rangle)\right) ; w \in \mathbb{R}^k\right\}$$

مجموعه $\phi(x_1) \dots \phi(x_d)$ در فضای $\mathbb{R}^k$ وابسته خطی خواهند بود و بنابراین shatter نخواهند شد

⟵ بنابراین نشان دادیم که d سمپل n بعدی را می‌توان به فضای d بعدی برده و shatter کرد اما به محض آنکه تعداد سمپل ها از بعد فضا (یعنی d) بیشتر شود دیگر نمی‌توان آنها را shatter کرد

$$d_{VC}(H) = dim(F) = d$$

در کوئرا اشاره شد که منظور سوال آن بوده که توابع F خطی اند. اگر این طور باشد نیز می‌دانیم که جمع 2 تابع حقیقی و خطی، حقیقی می‌شود و مانند قبل که نشان دادیم جواب معادل $d_{VC}(\mathrm{sign}(f))$ است. این بار می‌توان برای $d_{VC}(\mathrm{sign}(g))$ اثبات کرد که دقیقا مطابق اثبات ذکر شده می‌شود و به جواب نهایی d می‌رسیم

$\boxed{2.d}$

می‌خواهیم نشان دهیم که $d_{VC}(H) = 3$

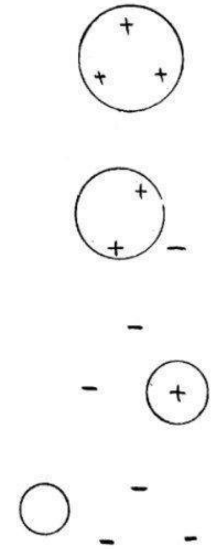
$H \rightarrow$ circles in $\mathbb{R}^2$

lower bound

ابتدا نشان می‌دهیم مثالی وجود دارد که 3 سمپل را در فضای $\mathbb{R}^2$ بچینیم و H بتواند آن را shatter کند

یعنی به ازای تمام $8 = 2^3$ حالت ممکن برای لیبل‌ها، H بتواند سمپل‌ها را دسته‌بندی کند

به هرصورت که نقاط را در feature space قرار دهیم، حالات زیر متصور است



- هر 3 نقطه لیبل + — 1 حالت
- دو نقطه لیبل + — 3 حالت
- یک نقطه لیبل + — 3 حالت
- تمام نقاط لیبل − — 1 حالت

هر 8 حالت دسته‌بندی شد ← $d_{VC}(H) \geq 3$ (I)

upper bound

حال باید نشان دهیم که اگر 4 سمپل را در فضای $\mathbb{R}^2$ بچینیم، به هر نحوی که این چیدمان را انجام دهیم، H نمی‌تواند تمام $2^4=16$ حالت ممکن برای لیبل‌ها را دسته‌بندی کند. (این 4 نقطه را x_1، x_2، x_3، x_4 می‌نامیم)

اگر این 4 نقطه روی مرز یک convex hull قرار نداشته باشند، از همان ابتدا معلوم است که حداقل یکی از آن‌ها بین دیگران قرار می‌گیرد و به این صورت دایره‌ای وجود نخواهد داشت که این نقطه را مخالف 3 نقطه دیگر لیبل بزند. پس فرض می‌کنیم که 4 نقطه روی مرز یک convex hull قرار دارند. حال فرض می‌کنیم که ترتیب قرار گرفتن نقاط روی convex hull مطابق شکل رسم شده باشد.

می‌دانیم که در چهارضلعی تشکیل شده حتماً یک جفت از زوایای روبه‌رو جمعشان کمتر مساوی $180°$ است (در این شکل $\hat{x}_1 + \hat{x}_3 \leq 180°$)

حال می‌خواهیم ثابت کنیم که هیچ دایره‌ای نخواهیم داشت که x_1 و x_3 را شامل شود اما هیچ کدام x_2 یا x_4 را شامل نشود. فرض کنیم x_2 بخواهد خارج این دایره باشد پس بنابراین کمان دایره که روبه‌روی x_2 است و بین x_1 و x_3 قرار دارد زاویه‌ای کمتر از $180 - \hat{x}_2$ خواهد داشت.

در حقیقت تمام نقاط بیرون این دایره، زاویه‌ای که با x_1 و x_3 می‌سازند کمتر از $180 - \hat{x}_2$ خواهد بود و تمام نقاط داخل دایره، زاویه‌ای که با x_1 و x_3 می‌سازند بیشتر از $180 - \hat{x}_2$ خواهد بود. از آن جا که $\hat{x}_1 + \hat{x}_3 \leq 180°$ پس $\hat{x}_2 + \hat{x}_4 \geq 180°$ ← $x_4 \geq 180 - \hat{x}_2$

بنابراین x_4 داخل دایره قرار می‌گیرد. به همین صورت نیز اگر x_4 بخواهد خارج دایره باشد، x_2 داخل آن قرار می‌گیرد.

پس نشان دادیم که امکان ندارد x_1 و x_3 را یک لیبل بزنیم و x_2 و x_4 را لیبلی مخالف آن ← $d_{VC}(H) < 4$ (II)

$\text{(I)(II)} \Longrightarrow \boxed{d_{VC}(H) = 3}$

پرسش ۲

$\boxed{1}$ $e(h_0) \leqslant e(h^*) + \eta \longrightarrow p\left[e(h_0) - e(h^*) \leqslant \eta\right] = 1$ ①

$\delta = 2ke^{-2m\gamma^2}$

طبق هوفدینگ $p\left[e(h^*) - \hat{e}(h^*) < \gamma\right] > 1-\delta$ ②

$\xrightarrow{①+②} p\left[e(h_0) - \hat{e}(h^*) < \eta + \gamma\right] > 1-\delta$

$\longrightarrow p\left[e(h_0) - \hat{e}(h^*) > \eta + \gamma\right] \leqslant \delta$

می دانیم که از نظر $\hat{e}$ کمترین ارور برای $\hat{h}$ است پس می توانیم به جای $\hat{e}(h^*)$ ، $\hat{e}(\hat{h})$ را جایگزین کنیم و همچنان نامعادله برقرار خواهد بود. همچنین طبق هوفدینگ

$p\left[\hat{e}(h_0) - e(h_0) \geqslant \gamma\right] \leqslant \delta$ یعنی اگر به جای $e(h_0)$ ، $\hat{e}(h_0)$ را قرار دهیم یک γ اضافه می شود و چون سمت راست بزرگتر می شود نامساوی همچنان برقرار خواهد بود

$p\left[\hat{e}(h_0) - \gamma - \hat{e}(\hat{h}) > \eta + \gamma\right] \leqslant \delta$

$\longrightarrow p\left[\hat{e}(h_0) > \hat{e}(\hat{h}) + \eta + 2\gamma\right] \leqslant \delta \longrightarrow$ احتمال NO برگرداندن حداکثر δ است

$\boxed{2}$

$$e(h_0) > e(h^*) + \eta \longrightarrow P\left[e(h_0) - e(h^*) > \eta\right] = 1 \quad (1)$$

طبق هوفدینگ $\quad P\left[\hat{e}(h^*) - e(h^*) < \gamma\right] > 1-\delta \longrightarrow P\left[e(h^*) - \hat{e}(h^*) > -\gamma\right] > 1-\delta \quad (2)$

$$\xrightarrow{(1)+(2)} P\left[e(h_0) - \hat{e}(h^*) > \eta - \gamma\right] > 1-\delta$$

مطابق قبل می‌توانیم به‌جای $\hat{e}(h^*)$ عبارت $\hat{e}(\hat{h})$ که کوچکتر است را قرار دهیم و همچنان نامعادله برقرار خواهد بود

همچنین طبق هوفدینگ

$$P\left[e(h_0) - \hat{e}(h_0) < \gamma\right] > 1-\delta$$

$$P\left[\hat{e}(h_0) - e(h_0) > -\gamma\right] > 1-\delta$$

یعنی اگر به‌جای $e(h_0)$، $\hat{e}(h_0)$ را قرار دهیم یک $-\gamma$ اضافه می‌شود و نامساوی همچنان برقرار است

$$P\left[\hat{e}(h_0) - \hat{e}(\hat{h}) > \eta - 2\gamma\right] > 1-\delta$$

$$\longrightarrow P\left[\hat{e}(h_0) < \hat{e}(\hat{h}) + \eta - 2\gamma\right] \leq \delta$$

احتمال yes برگرداندن حداکثر δ است

$\boxed{3}$

طبق هوفدینگ $\quad P\left[\hat{e}(h^*) - e(h^*) < \gamma\right] > 1-\delta$

می‌دانیم که از نظر e کمترین ارور برای h^* است پس می‌توانیم به‌جای $e(h^*)$، $e(\hat{h})$ را جایگذین کنیم زیرا قطعاً بزرگتر یا مساوی آن است و نامعادله برقرار می‌ماند

$$P\left[\hat{e}(h^*) - e(\hat{h}) < \gamma\right] > 1-\delta \quad (1)$$

طبق هوفدینگ $\quad P\left[e(\hat{h}) - \hat{e}(\hat{h}) < \gamma\right] > 1-\delta \quad (2)$

$$\xrightarrow{(1)+(2)} P\left[\hat{e}(h^*) - \hat{e}(\hat{h}) < 2\gamma\right] > 1-\delta$$

$$\xrightarrow[h^* = h_0]{2\gamma < \eta - 2\gamma \quad \text{به ازای m های بزرگ}} P\left[\hat{e}(h_0) < \hat{e}(\hat{h}) + \eta - 2\gamma\right] > 1-\delta$$

احتمال yes برگرداندن حداقل $1-\delta$ است

پرسش ۳

$\boxed{a}$

$$\frac{d\,\phi(1-\phi)}{d\phi} = -2\phi+1=0 \longrightarrow \phi=\frac{1}{2} \longrightarrow \max\left(var(z_i)\right)=\frac{1}{4}$$

$\boxed{b}$

اگر متغیر تصادفی $z_i' \sim Bernoulli(e(\hat{h}))$ را در نظر بگیریم

$\boxed{1}$

$$\xrightarrow[\text{جاگذاری حداکثر واریانس}]{\text{طبق نابرابری چبیشف}} P\left(|\hat{e}(\hat{h}) - e(\hat{h})|>\gamma\right) < \frac{Var(z_i')}{m\gamma^2} = \frac{e(\hat{h})\left(1-e(\hat{h})\right)}{m\gamma^2} \leqslant \frac{1}{4m\gamma^2}$$

$\boxed{2}$

$$P\left(|\hat{e}(\hat{h}) - e(\hat{h})|>\gamma\right) < \underbrace{\frac{1}{4m\gamma^2}\leqslant\delta}_{\circledast}$$

$$\xrightarrow{\circledast} 4m\gamma^2 \geqslant \frac{1}{\delta} \longrightarrow m\geqslant \frac{1}{4\delta\gamma^2}$$

$\circledast$ حال اگر $\delta\in(0,1)$ فرض کنیم برای m داریم:

$$m\in\left[\frac{1}{4\gamma^2},\infty\right)$$

$\boxed{3}$

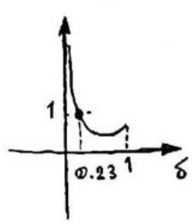
$$\xrightarrow[\text{استفاده کنیم}]{\text{اگر از هوفدینگ}} 2e^{-2\gamma^2 m}\leqslant\delta \longrightarrow m\geqslant \frac{-\ln\frac{\delta}{2}}{2\gamma^2}$$

$\circledast$ حال اگر $\delta\in(0,1)$ فرض کنیم برای m داریم

$$m\in\left[\frac{\ln 2}{2\gamma^2},\infty\right)$$

برای مقایسه حداقل m هایی که از چبیشف به دست می‌آید با حداقل m هایی که از هوفدینگ به دست می‌آید، آن 2 را برهم تقسیم می‌کنیم

$$\frac{\frac{1}{4\delta\gamma^2}}{\frac{-\ln\frac{\delta}{2}}{2\gamma^2}} = \frac{1}{-2\delta\ln\frac{\delta}{2}}$$



نمودار حاصل تقسیم را برای $\delta\in(0,1)$ رسم کردیم و مشخص است که تقریبا

در بازه $\delta\in(0,0.23)$ چبیشف m بیشتری اعلام می‌کند ← هوفدینگ بهتر است

در بازه $\delta\in(0.23,1)$ هوفدینگ m بیشتری اعلام می‌کند ← چبیشف بهتر است

$\boxed{C}$

برای k تا فرضیه داریم :

$$P\left(|e(h_1)-\hat{e}(h_1)|>\gamma \cup \cdots \cup |e(h_k)-\hat{e}(h_k)|>\gamma\right) \leqslant \sum_{i=1}^{k} P\left(|e(h_i)-\hat{e}(h_i)|>\gamma\right)$$

$$< \underbrace{\frac{k}{4m\gamma^2}}_{\text{برابر } \delta \text{ می‌گیریم}}$$

$$\longrightarrow \delta = \frac{k}{4m\gamma^2} \longrightarrow \gamma = \frac{1}{2}\sqrt{\frac{k}{m\delta}}$$

اگر متمم احتمال اولیه را در نظر بگیریم، داریم که

$$P\left(|e(\hat{h})-\hat{e}(\hat{h})|<\gamma\right) > 1-\delta$$

بنابراین generalization error بهترین فرضیه با حداقل احتمال $1-\delta$ به γ و بنابراین $\sqrt{\frac{k}{m\delta}}$ محدود می‌شود