دانشکده علوم ریاضی
دانشگاه صنعتی شریف

# نظریه یادگیری محاسباتی

**امید اعتصامی، محمدهادی فروغمنداعرابی**

بهار ۱۳۹۳

# بعد VC

جلسه‌های ؟

**نگارنده:**

## ۱ مقدمه

در جلسات پیش، با استفاده از تیغ اُکام، دیدیم که کلاس‌های متناهی همواره قابل یادگیری PAC هستند. همچنین مشاهده کردیم که کلاس نامتناهی مستطیل‌های با اضلاع موازی محورها، قابل یادگیری PAC است. بنابراین طبیعی است که بپرسیم چه کلاس‌هایی قابل یادگیری PAC هستند. در این بخش یک مفهوم بُعد از نوع ترکیبیاتی معرفی می‌کنیم و نشان می‌دهیم یک کلاس قابل یادگیری PAC است اگر و تنها اگر بُعد آن متناهی باشد.

در این فصل در مورد کارایی الگوریتم یادگیری هیچ صحبتی نمی‌کنیم و تنها به قابلیت یادگیری یک کلاس مفهوم می‌پردازیم.

## ۲ بعد VC: تعریف و چند مثال

ایده‌ی تعریف بُعد، قضیه‌ی "ناهار مجانی در کار نیست!" می‌باشد. این قضیه بیان می‌کند که اگر هیچ محدودیتی روی کلاس مفهوم وجود نداشته باشد، آنگاه آن کلاس قابل یادگیری نیست. در واقع اگر کلاس مفهوم شامل همه‌ی توابع ممکن(همه‌ی توابع $f : X \to \{0,1\}$) باشد، آنگاه قابل یادگیری PAC نیست.

**تعریف ۱.** *گیریم $C = \{c_1, c_2, \cdots, c_m\} \subseteq X$ و $\mathcal{H}$ کلاسی از توابع $f : X \to \{0,1\}$ باشد. تحدید $\mathcal{H}$ به $C$ به صورت زیر تعریف می‌شود*

$$\mathcal{H}_C = \{(h(c_1), \cdots, h(c_m)) : h \in \mathcal{H}\}$$

**تعریف ۲.** *یک کلاس فرض $\mathcal{H}$، زیر مجموعه‌ی متناهی $C \subseteq X$ را خرد می‌کند، هرگاه $|\mathcal{H}_C| = 2^{|C|}$. یعنی تحدید $\mathcal{H}$ به $C$، همه‌ی توابع ممکن از $C$ به $\{0,1\}$ را تشکیل دهد.*

**مثال ۳.** *فرض کنید $\mathcal{H} = \{h_\theta : \theta \in \mathbb{R}\}$ که در آن $h_\theta(x) = \begin{cases} 0 & x < \theta \\ 1 & x \geq \theta \end{cases}$ هرگاه $x \in \mathbb{R}$ آنگاه $h_x(x) = 1$ و $h_{x+1}(x) = 0$. بنابراین $\mathcal{H}$ هر مجموعه‌ی تک عضوی را خرد می‌کند. حال فرض کنیم $C = \{x, y\}$ یک مجموعه‌ی دو عضوی باشد. در این صورت نشان می‌دهیم $\mathcal{H}$ نمی‌تواند $C$ را خرد کند. بدون از دست دادن کلیت فرض کنیم $x < y$. در این حالت عنصر $(1,0)$ هرگز تولید نخواهد شد زیرا اگر $h_\theta$ تابعی باشد که $h_\theta(y) = 0$ آنگاه $x < y < \theta$ لذا $h_\theta(x) = 0$. در نتیجه $\mathcal{H}$ هیچ مجموعه‌ی دو عضوی را خرد نمی‌کند.*

**مثال ۴.** *فرض کنید $\mathcal{H} = \{h_{\theta,\theta'} : \theta, \theta' \in \mathbb{R}, \theta \leq \theta'\}$ که در آن $h_\theta(x) = \begin{cases} 0 & \theta \leq x \leq \theta' \\ 1 & o.w. \end{cases}$ در این صورت به ازای هر $x, y \in \mathbb{R}$ با ویژگی $x < y$، $(h_{x,y}(x), h_{x,y}(y)) = (1,1)$ اگر $r > y$ آنگاه $(h_{r,r+1}(x), h_{r,r+1}(y)) = (0,0)$. همچنین اگر $r = \frac{y-x}{2}$ آنگاه $(h_{x-r,x+r}(x), h_{x-r,x+r}(y)) = (1,0)$ و $(h_{y-r,y+r}(x), h_{y-r,y+r}(y)) = (0,1)$. لذا $\mathcal{H}$، هر مجموعه‌ی دو عضوی را خرد می‌کند. حال نشان می‌دهیم $\mathcal{H}$ هیچ مجموعه‌ی سه عضوی را خرد نمی‌کند. هرگاه $x$، $y$ و $z$ سه عدد حقیقی باشند که $x < y < z$ آنگاه اگر $h_{\theta,\theta'}(x) = h_{\theta,\theta'}(z) = 1$ آنگاه $h_{\theta,\theta'}(y) = 1$. بنابراین عنصر $(1,0,1)$ هرگز تولید نمی‌شود.*

حال بُعد VC را تعریف می‌کنیم.

**تعریف ۵.** *تعداد اعضای بزرگ‌ترین مجموعه‌ی متناهی $C \subseteq X$ که توسط $\mathcal{H}$ خرد می‌شود، در صورت وجود بیشینه، بعد VC نامیده می‌شود. اگر این مقدار کران‌دار نباشد قرار می‌دهیم $VCdim(\mathcal{H}) = \infty$.*

**مثال ۶.** *برای کلاس مستطیل‌های با اضلاع موازی محورهای مختصات داریم $VCdim = 4$ (چرا؟).*

**قضیه ۷.** *کلاس‌های با بعد VC نامتناهی قابل یادگیری PAC نیستند.*

*اثبات.* اثبات نتیجه‌ای مستقیم از قضیه‌ی "ناهار مجانی در کار نیست!" می‌باشد. □

**قضیه ۸.** *هرگاه $\mathcal{H}$ کلاس مفهومی متناهی باشد آنگاه $VCdim(\mathcal{H}) \leq log_2(|\mathcal{H}|)$.*

*اثبات.* هرگاه $C$ مجموعه‌ای باشد که توسط $\mathcal{H}$ خرد می‌شود آنگاه $2^{|C|} = |\mathcal{H}_C| \leq |\mathcal{H}|$. در نتیجه $|C| \leq log_2(|\mathcal{H}|)$ و لذا $VCdim(\mathcal{H}) \leq log_2(|\mathcal{H}|)$. □

**تذکر ۹.** *ممکن است در مثال‌ها مشاهده کرده باشید که بعد VC با تعداد پارامترها برابر بود. این حکم در حالت کلی درست نیست. در واقع اگر $\mathcal{X} = \mathbb{R}$ و $\mathcal{H} = \{h_\theta : \theta \in \mathbb{R}\}$ که در آن $h_\theta(x) = \lceil 0.5 sin(\theta x) \rceil$ آنگاه این مجموعه دارای یک پارامتر است اما بعد VC آن $\infty$ است.*

# ۳ قضیه‌ی اساسی یادگیری آماری

در این بخش نشان می‌دهیم یک کلاس قابل یادگیری PAC است اگر و تنها اگر بعد VC آن متناهی باشد. ایده‌ی کار آن است که تحدید کلاس‌های با بعد متناهی روی مجموعه‌ی C خیلی آهسته رشد می‌کند و با این ویژگی می‌توان همگرایی یکنواخت را ثابت کرد. از قبل می‌دانیم همگرایی یکنواخت یادگیری را تضمین می‌کند.

**تعریف ۱۰.** *گیریم $\mathcal{H}$ یک کلاس فرض باشد. تابع رشد $\mathcal{H}$، با نماد $\tau_{\mathcal{H}} : \mathbb{N} \to \mathbb{N}$ به صورت زیر تعریف می‌شود*

$$\tau_{\mathcal{H}}(m) = \max_{C \subseteq \mathcal{X} : |C| = m} |\mathcal{H}_C| \quad (۱)$$

**لم ۱۱** (Sauer-Shllah-Perles)**.** *گیریم $\mathcal{H}$ یک کلاس فرض با $VCdim(\mathcal{H}) \le d < \infty$ باشد. در این صورت، برای هر $m$،*

$$\tau_{\mathcal{H}}(m) \le \sum_{i=0}^{d} \binom{m}{i} \quad (۲)$$

*به ویژه اگر $m > d + 1$ آنگاه $\tau_{\mathcal{H}}(m) \le (em/d)^d$.*

*اثبات.* به جای اثبات حکم اصلی، حکمی قوی‌تر ثابت می‌کنیم: برای هر $C = \{c_1, c_2, \cdots, c_m\}$ نشان می‌دهیم

$$\forall \mathcal{H}, \quad |\mathcal{H}_C| \le |\{B \subseteq C : \mathcal{H} \text{ shatters } B\}| \quad (۳)$$

ابتدا ببینیم چرا این حکم، حکم اصلی را نتیجه می‌دهد.

چون $VCdim(\mathcal{H}) \le d$ پس $\mathcal{H}$ هیچ مجموعه‌ی با تعداد عضو بیشتر از d را خرد نمی‌کند. در نتیجه

$$|\{B \subseteq C : \mathcal{H} \text{ shatters } B\}| \le \sum_{i=0}^{d} \binom{m}{i} \quad (۴)$$

وقتی که $m \ge d + 1$ داریم $\frac{d}{m} < 1$ بنابراین

$$(\frac{d}{m})^d \sum_{i=0}^{d} \binom{m}{i} \le \sum_{i=0}^{d} (\frac{d}{m})^i \binom{m}{i} \le \sum_{i=0}^{m} (\frac{d}{m})^i \binom{m}{i} = (1 + \frac{d}{m})^m \le e^d \quad (۵)$$

با تقسیم طرفین بر $(\frac{d}{m})^d$ قسمت دوم حکم نیز به دست می‌آید.

حال به اثبات ۳ می‌پردازیم. برای $m = 1$ هر دو طرف برابر ۱ یا ۲ هستند. فرض کنیم ۳ برای هر $k < m$ برقرار باشد. حکم را برای m ثابت خواهیم کرد. $\mathcal{H}$ و $C = \{c_1, c_2, \cdots, c_m\}$ را ثابت بگیرید. قرار می‌دهیم $C\prime = \{c_2, \cdots, c_m\}$ و دو مجموعه‌ی

$$Y_0 = \{(y_2, \cdots, y_m) : (0, y_2, \cdots, y_m) \in \mathcal{H}_C \vee (1, y_2, \cdots, y_m) \in \mathcal{H}_C\} \quad (۶)$$

و

$$Y_1 = \{(y_2, \cdots, y_m) : (0, y_2, \cdots, y_m) \in \mathcal{H}_C \wedge (1, y_2, \cdots, y_m) \in \mathcal{H}_C\} \quad (۷)$$

را در نظر می‌گیریم. ابتدا نشان می‌دهیم $|\mathcal{H}_C| = |Y_0| + |Y_1|$. دو حالت وجود دارد:

- اگر فقط یکی از $(0, y_2, \cdots, y_m)$ و $(1, y_2, \cdots, y_m)$ عضو $\mathcal{H}_C$ باشد آنگاه این عنصر متعلق به $Y_1$ نیست و به هر طرف تساوی عدد ۱ اضافه می‌شود.

- اگر هر دو این عناصر در $\mathcal{H}_C$ وجود داشته باشند آنگاه به طرف چپ عدد ۲ اضافه می‌شود و به هر کدام از $|Y_0|$ و $|Y_1|$ عدد ۱ اضافه می‌شود. پس در کل به طرف راست عدد ۲ اضافه می‌شود.

این ادعا را ثابت می‌کند. از طرفی چون $Y_0 = \mathcal{H}_{C'}$ طبق فرض استقرا

$$|Y_0| = |\mathcal{H}_{C'}| \leq |\{B \subseteq C' : \mathcal{H} \ shatters \ B\}| = |\{B \subseteq C : c_1 \notin B \wedge \mathcal{H} \ shatters \ B\}| \tag{۸}$$

حال مجموعه‌ی $\mathcal{H}' \subseteq \mathcal{H}$ را به صورت زیر تعریف می‌کنیم

$$\mathcal{H}' = \{h \in \mathcal{H} : \exists h' \in \mathcal{H} \ s.t. \ (1 - h'(c_1), h'(c_2), \cdots, h'(c_m)) = (h(c_1), h(c_2), \cdots, h(c_m))\} \tag{۹}$$

به وضوح $Y_1 = \mathcal{H}'_{C'}$ و $\mathcal{H}'$ مجموعه‌ی $B \subseteq C'$ را خرد می‌کند اگر و تنها اگر $B \cup \{c_1\}$ را خرد کند. بنابراین با استفاده از فرض استقرا داریم

$$\begin{aligned} |Y_0| = |\mathcal{H}_{C'}| &\leq |\{B \subseteq C' : \mathcal{H}' \ shatters \ B\}| \\ &= |\{B \subseteq C' : \mathcal{H}' \ shatters \ B \cup \{c_1\}\}| \\ &= |\{B \subseteq C : c_1 \in B \wedge \mathcal{H}' \ shatters \ B\}| \\ &\leq |\{B \subseteq C : c_1 \in B \wedge \mathcal{H} \ shatters \ B\}| \end{aligned} \tag{۱۰}$$

بنابراین در کل داریم

$$\begin{aligned} |\mathcal{H}_C| &= |Y_0| + |Y_1| \\ &\leq |\{B \subseteq C : c_1 \notin B \wedge \mathcal{H} \ shatters \ B\}| + |\{B \subseteq C : c_1 \in B \wedge \mathcal{H} \ shatters \ B\}| \\ &= |\{B \subseteq C : \mathcal{H} \ shatters \ B\}| \end{aligned} \tag{۱۱}$$

این، حکم را نتیجه می‌دهد. □

**قضیه ۱۲.** *گیریم $\mathcal{H}$ یک کلاس و $\tau_{\mathcal{H}}(m)$ تابع رشد باشد. در این صورت برای هر $\delta \in (0,1)$ با احتمال حداقل $1-\delta$ روی انتخاب $S \sim D^m$ داریم*

$$|L_{\mathcal{D}}(h) - L_S(h)| \leq \frac{4 + \sqrt{\log(\tau_{\mathcal{H}}(2m))}}{\delta\sqrt{2m}} \tag{۱۲}$$

*اثبات.* کافی است نشان دهیم

$$\mathop{\mathbb{E}}_{S\sim\mathcal{D}^m}(\sup_{h\in\mathcal{H}} |L_{\mathcal{D}}(h) - L_S(h)|) \leq \frac{4 + \sqrt{\log(\tau_{\mathcal{H}}(2m))}}{\delta\sqrt{2m}} \tag{۱۳}$$

در این صورت با استفاده از نامساوی مارکوف حکم به سادگی حاصل می‌شود.

برای کران کردن سمت چپ معادله‌ی ۱۳ ابتدا توجه می‌کنیم که $L_{\mathcal{D}}(h) = \mathbb{E}_{S'\sim\mathcal{D}^m}(L_{S'}(h))$ که در آن $S' = z'_1, \cdots, z'_m$ یک نمونه‌ی مستقل دیگر است. بنابراین

$$\begin{aligned} \mathop{\mathbb{E}}_{S\sim\mathcal{D}^m}(\sup_{h\in\mathcal{H}} |L_{\mathcal{D}}(h) - L_S(h)|) &= \mathop{\mathbb{E}}_{S\sim\mathcal{D}^m}(\sup_{h\in\mathcal{H}} |\mathop{\mathbb{E}}_{S'\sim\mathcal{D}^m} L_{S'}(h) - L_S(h)|) \\ &\leq \mathop{\mathbb{E}}_{S\sim\mathcal{D}^m}(\sup_{h\in\mathcal{H}} \mathop{\mathbb{E}}_{S'\sim\mathcal{D}^m} |L_{S'}(h) - L_S(h)|) \\ &\leq \mathop{\mathbb{E}}_{S\sim\mathcal{D}^m} \mathop{\mathbb{E}}_{S'\sim\mathcal{D}^m}(\sup_{h\in\mathcal{H}} |L_{S'}(h) - L_S(h)|) \\ &= \mathop{\mathbb{E}}_{S,S'\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{H}} \frac{1}{m} |\sum_{i=1}^{m}(l(h,z'_i) - l(h,z_i))|\right] \end{aligned} \tag{۱۴}$$

که در آن نامساوی خط دوم از نامساوی مثلثی و نامساوی خط سوم نیز به دلیل امکان تعویض امید ریاضی و sup حاصل شده است.

در آخرین تساوی معادلات ۱۴، به دلیل مستقل بودن نمونه‌ها می‌توان نام آن‌ها را با هم عوض کرد. بنابراین برای هر $\sigma \in \{-1,1\}^m$،

$$\begin{aligned}\mathop{\mathbb{E}}_{S,S'\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{H}}\frac{1}{m}\Big|\sum_{i=1}^{m}(l(h,z_i')-l(h,z_i))\Big|\right] &= \mathop{\mathbb{E}}_{S,S'\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{H}}\frac{1}{m}\Big|\sum_{i=1}^{m}\sigma_i(l(h,z_i')-l(h,z_i))\Big|\right]\\ &= \mathop{\mathbb{E}}_{\sigma\sim U^m}\mathop{\mathbb{E}}_{S,S'\sim\mathcal{D}^m}\left[\sup_{h\in\mathcal{H}}\frac{1}{m}\Big|\sum_{i=1}^{m}\sigma_i(l(h,z_i')-l(h,z_i))\Big|\right]\\ &= \mathop{\mathbb{E}}_{S,S'\sim\mathcal{D}^m}\mathop{\mathbb{E}}_{\sigma\sim U^m}\left[\sup_{h\in\mathcal{H}}\frac{1}{m}\Big|\sum_{i=1}^{m}\sigma_i(l(h,z_i')-l(h,z_i))\Big|\right]\end{aligned} \tag{۱۵}$$

که منظور از $U$، توزیع یکنواخت روی $\{-1,1\}$ است. تساوی دوم به این دلیل است که وقتی برای هر $\sigma$ تساوی برقرار است، پس برای امید یکنواخت نیز برقرار است. در واقع میانگین تعدادی عدد مساوی با آن اعداد مساوی است. تساوی آخر نیز به دلیل خطی بودن امید ریاضی است. بنابراین برای اثبات حکم کافی است نامساوی مطلوب را برای آخرین امید ریاضی معادله‌ی ۱۵ ثابت کنیم.

در اینجا حکمی کلی‌تر ثابت می‌کنیم یعنی نشان می‌دهیم برای هر $S$ و $S'$،

$$\mathop{\mathbb{E}}_{\sigma\sim U^m}\left[\sup_{h\in\mathcal{H}}\frac{1}{m}\Big|\sum_{i=1}^{m}\sigma_i(l(h,z_i')-l(h,z_i))\Big|\right] \tag{۱۶}$$

از مقدار مطلوب کمتر است(هرگاه تعدادی عدد از یک عدد کمتر باشند، میانگین آن‌ها نیز از آن عدد کمتر است). حال که $S$ و $S'$ ثابت شده‌اند، می‌توان $\mathcal{H}$ را به $C = S\cup S'$ تحدید کرد. با این کار می‌توان sup را روی $\mathcal{H}_C$ گرفت و به خاطر متناهی بودن $\mathcal{H}_C$ می‌توان sup را به max تبدیل نمود. بنابراین داریم

$$\mathop{\mathbb{E}}_{\sigma\sim U^m}\left[\sup_{h\in\mathcal{H}}\frac{1}{m}\Big|\sum_{i=1}^{m}\sigma_i(l(h,z_i')-l(h,z_i))\Big|\right] = \mathop{\mathbb{E}}_{\sigma\sim U^m}\left[\max_{h\in\mathcal{H}_C}\frac{1}{m}\Big|\sum_{i=1}^{m}\sigma_i(l(h,z_i')-l(h,z_i))\Big|\right] \tag{۱۷}$$

حال برای هر $h\in\mathcal{H}_C$ قرار می‌دهیم $\theta_h = \frac{1}{m}\sum_{i=1}^{m}\sigma_i(l(h,z_i')-l(h,z_i))$ در نتیجه کافی است نامساوی را برای $\mathop{\mathbb{E}}_{\sigma\sim U^m}[\max_{h\in\mathcal{H}_C}|\theta_h|]$ ثابت کنیم. در ادامه همه‌ی احتمال‌ها و امیدها روی $\sigma\sim U^m$ است که برای راحتی آن را نمی‌نویسیم.

چون $\mathbb{E}(\theta_h)=0$ و $\theta\in[-1,1]$، با استفاده از نامساوی Hoeffding، برای هر $\rho>0$ داریم

$$\begin{aligned}\Pr[|\theta_h|>\rho] \le 2\exp\left(-2m\rho^2\right) &\Rightarrow \Pr\left[\max_{h\in\mathcal{H}_C}|\theta_h|>\rho\right] \le 2|\mathcal{H}_C|\exp\left(-2m\rho^2\right)\\ &\Rightarrow \mathop{\mathbb{E}}_{\sigma\sim U^m}\left[\max_{h\in\mathcal{H}_C}|\theta_h|\right] \le \frac{4+\sqrt{\log(\tau_{\mathcal{H}}(2m))}}{\delta\sqrt{2m}}\end{aligned} \tag{۱۸}$$

در معادلات بالا، اولین نتیجه گیری از union bound و دومین نتیجه‌گیری از از لم A.4 در ضمیمه‌ی A کتاب، به دست آمده است. این لم را، بدون اثبات، در ادامه بیان می‌کنیم. نتیجه‌ی آخر همان نامساوی مطلوب است. بدین ترتیب حکم به اثبات می‌رسد. $\square$

**لم ۱۳.** *گیریم $X$ یک متغیر تصادفی، $x'\in\mathbb{R}$ یک اسکالر و $a>0$ و $b\ge e$ وجود داشته باشد که برای هر $t\ge 0$ داشته باشیم $\Pr[|X-x'|>t]\le 2b\, e^{-(t^2/a^2)}$ در این صورت $\mathbb{E}[|X-x'|]\le a(2+\sqrt{\log(b)})$.*

**قضیه ۱۴** (قضیه‌ی اساسی یادگیری آماری نسخه‌ی کیفی). *گیریم $\mathcal{H}$ یک کلاس فرض از توابع $\mathcal{X}\to\{0,1\}$ و تابع هزینه*[1] *یک تابع هزینه‌ی $1-0$ باشد. آنگاه گزاره‌های زیر معادلند*

[1] Loss Function

۱. *کلاس $\mathcal{H}$ دارای ویژگی همگرایی یکنواخت است.*

۲. *هر الگوی ERM یک یادگیر Agn. PAC برای $\mathcal{H}$ است.*

۳. *کلاس $\mathcal{H}$ قابل یادگیری Agn. PAC است.*

۴. *کلاس $\mathcal{H}$ قابل یادگیری PAC است.*

۵. *هر الگوی ERM یک یادگیر PAC برای $\mathcal{H}$ است.*

۶. *کلاس $\mathcal{H}$ دارای بعد VC متناهی است.*

*اثبات.* ما در این جا $۱ \Rightarrow ۶$ را ثابت می‌کنیم. مابقی احکام از نتایج بخش‌های قبل نتیجه می‌شود. برای این نیز ثابت می‌کنیم

$$m_{\mathcal{H}}^{UC} \leq 4\frac{16d}{(\delta\epsilon)^2}\log(\frac{16d}{(\delta\epsilon)^2}) + \frac{16d\log(2e/d)}{(\delta\epsilon)^2} \tag{19}$$

طبق لم Sauer برای $m > d$ داریم $\tau_{\mathcal{H}}(2m) \leq (2em/d)^d$. از ترکیب این با قضیه‌ی ۱۲ می‌توان نشان داد که با احتمال حداقل $1-\delta$ داریم

$$|L_S(h) - L_{\mathcal{D}}(h)| \leq \frac{4+\sqrt{d\log(2em/d)}}{\delta\sqrt{2m}} \tag{20}$$

برای سادگی فرض کنیم $\sqrt{d\log(2em/d)} \geq 4$ از این رو

$$|L_S(h) - L_{\mathcal{D}}(h)| \leq \frac{1}{\delta}\sqrt{\frac{2d\log(2em/d)}{m}} \tag{21}$$

برای این که مطمئن باشیم خطا از $\epsilon$ کمتر است، نیاز داریم که

$$m \geq 4\frac{2d}{(\delta\epsilon)^2}\log(\frac{2d}{(\delta\epsilon)^2}) + \frac{4d\log(2e/d)}{(\delta\epsilon)^2} \tag{22}$$

این قضیه را اثبات می‌کند. □

**قضیه ۱۵** (قضیه‌ی اساسی یادگیری آماری نسخه‌ی کمی). *گیریم $\mathcal{H}$ یک کلاس فرض از توابع $\mathcal{X} \to \{0,1\}$ و تابع هزینه*[۲] *یک تابع هزینه‌ی ۰−۱ باشد. فرض کنید $VCdim(\mathcal{H}) = d < \infty$ در این صورت ثابت‌های $C_1$ و $C_2$ وجود دارند به طوری که*

۱. *کلاس $\mathcal{H}$ دارای ویژگی همگرایی یکنواخت با پیچیدگی نمونه*[۳]

$$C_1\frac{d+\log(\frac{1}{\delta})}{\epsilon^2} \leq m_{\mathcal{H}}^{UC}(\epsilon,\delta) \leq C_2\frac{d+\log(\frac{1}{\delta})}{\epsilon^2} \tag{23}$$

*است.*

۲. *کلاس $\mathcal{H}$ قابل یادگیری Agn. PAC با پیچیدگی نمونه*

$$C_1\frac{d+\log(\frac{1}{\delta})}{\epsilon^2} \leq m_{\mathcal{H}}(\epsilon,\delta) \leq C_2\frac{d+\log(\frac{1}{\delta})}{\epsilon^2} \tag{24}$$

*است.*

---
[۲] Loss Function
[۳] Sample Complexity

۳. *کلاس $\mathcal{H}$ قابل یادگیری PAC با پیچیدگی نمونه*

$$C_1\frac{d+\log(\frac{1}{\delta})}{\epsilon} \leq m_{\mathcal{H}}(\epsilon,\delta) \leq C_2\frac{d\log(\frac{1}{\epsilon})+\log(\frac{1}{\delta})}{\epsilon} \tag{25}$$

*است.*

*اثبات.* اثبات این قضیه در بخش‌های بعدی مطرح خواهد شد. □

**قضیه ۱۶.** *بعد VC کلاس نیم‌فضاهای همگن در $\mathbb{R}^d$ برابر d است.*

*اثبات.* فرض کنید $e=\{e_1,e_2,\cdots,e_d\}$ پایه‌ی استاندارد $\mathbb{R}^d$ و $w=(y_1,y_2,\cdots,y_d)$ یک برچسب گذاری باشد. در این صورت $\langle w,e_i\rangle=y_i$ بنابراین e توسط کلاس نیم فضاهای همگن خرد می‌شود.

حال فرض کنید $C=\{x_1,\cdots,x_d,x_{d+1}\}\subseteq\mathbb{R}^d$ دلخواه باشد. اعداد حقیقی $a_1,\cdots,a_d$ وجود دارند که $x_{d+1}=\sum_{i=1}^d a_ix_i$ به وضوح برچسب گذاری $(-1,\cdots,-1,a_1+\cdots+a_d)$ هیچ وقت رخ نمی‌دهد(چرا؟). □

**قضیه ۱۷.** *بعد VC کلاس نیم‌فضاهای غیرهمگن در $\mathbb{R}^d$ برابر d است.*

*اثبات.* مانند قضیه‌ی ۱۶ می‌توان نشان داد $\{0,e_1,e_2,\cdots,e_d\}$ توسط کلاس نیم فضاهای غیرهمگن خرد می‌شود. از طرفی با توجه به مطالب بخش‌های قبل می‌دانیم کلاس نیم فضاهای غیرهمگن در $\mathbb{R}^d$ قابل تبدیل به کلاس نیم فضاهای همگن $\mathbb{R}^{d+1}$ است. بنابراین اگر $d+2$ نقطه در $\mathbb{R}^d$ توسط کلاس نیم فضاهای غیرهمگن خرد شود آنگاه توسط کلاس نیم فضاهای همگن در $\mathbb{R}^{d+1}$ خرد می‌شود که با قضیه‌ی ۱۶ در تناقض است. □