مسئله‌ی ۱.

فرض کنید دنباله‌ی توزیع $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_m$ روی دامنه‌ی $\mathcal{X}$ تعریف شده باشند. همچنین $\mathcal{H}$ یک کلاس فرضیه‌ی متناهی دودویی روی $\mathcal{X}$ باشد و داشته باشیم $f \in \mathcal{H}$. فرض کنید که مجموعه نمونه‌ی S شامل m نمونه را داشته باشیم، طوری که این نمونه‌ها از هم مستقل هستند، اما از توزیع یکسانی نیامده‌اند. در واقع فرض کنید که نمونه‌ی iام از توزیع $\mathcal{D}_i$ آمده است و مقدار y_i برابر با $f(x_i)$ باشد. فرض کنید که $\overline{\mathcal{D}}_m$ توزیع میانگین توزیع‌های ذکر شده باشد، به طور دقیق $\overline{\mathcal{D}}_m = (\mathcal{D}_1 + \mathcal{D}_2 + \ldots + \mathcal{D}_m)/m$. برای یک $\epsilon \in (0,1)$ نشان دهید:

$$\mathbb{P}\left[\exists h \in \mathcal{H} \text{ s.t. } L_{\left(\overline{\mathcal{D}}_m, f\right)}(h) > \epsilon \ \ and \ \ L_{(S,f)}(h) = 0\right] \leqslant |\mathcal{H}| e^{-\epsilon m}$$

مسئله‌ی ۲.

نشان دهید که برای هر توزیع $\mathcal{D}$، دسته‌بند بهینه‌ی بیز که آن را با $f_{\mathcal{D}}$ نشان می‌دهیم بهینه است، به این معنا که برای هر دسته‌بند $g : \mathcal{X} \to 0,1$ داشته باشیم:

$$L_{\mathcal{D}}(f_{\mathcal{D}}) \leqslant L_{\mathcal{D}}(g)$$

راهنمایی: در تعریف دسته بند بهینه بیز داریم، برای هر توزیع احتمالاتی $\mathcal{D}$ بر روی $\mathcal{X} \to 0,1$، بهترین تابع پیش‌بینی کننده برچسب از $\mathcal{X}$ به $0,1$ به شکل زیر خواهد بود:

$$f_{\mathcal{D}}(x) = \begin{cases} 1 & \text{if } \ \mathbb{P}[y=1 \mid x] \geqslant 1/2 \\ 0 & \text{otherwise} \end{cases}$$

مسئله‌ی ۳.

با فرض وجود الگوریتم یادگیری احتمالا تقریبا درست برای کلاس کانسپت $\mathcal{C}$ بر روی $\mathbb{R}^2$ که از تقاطع نیم فضاهای هم‌تراز با محور شکل گرفته‌است و کلاس کانسپت $\mathcal{C}$ از کانسپت‌های زیر تشکیل شده است:

$$\begin{aligned} &\{(x,y) : x \geqslant c_x, y \geqslant c_y\}, \\ &\{(x,y) : x \geqslant c_x, y \leqslant c_y\}, \\ &\{(x,y) : x \leqslant c_x, y \geqslant c_y\}, \\ &\{(x,y) : x \leqslant c_x, y \leqslant c_y\} \end{aligned}$$

به طوری که $c_x, c_y \in \mathbb{R}$. پیچیدگی نمونه را برای $\mathcal{C}$ بدست آوردید.

مسئله‌ی ۴.

یک نوع مدل PAC دو توزیعه وجود دارد که الگوریتم یادگیری آن ممکن است صراحتا نمونه های مثبت و منفی را درخواست کند اما باید فرضیه‌ای را بیابد که بر روی هر دو توزیع داده‌های مثبت و منفی به خوبی عمل کند.

زمانی می‌گوییم یک مدل PAC دو توزیعه با الگوریتم A به صورت احتمالا تقریبا صحیح کلاس فرضیه $\mathcal{H}$ را یاد می‌گیرد که اگر برای هر کانسپت هدف $c \in \mathcal{H}$ ، توزیع $\mathcal{D}^+$ بر روی داده‌های با برچسب مثبت، توزیع $\mathcal{D}^-$ بر روی داده‌های با برچسب منفی، برای هر (ϵ, δ) و اگر به اندازه کافی (اما محدود) داده مثبت و منفی به صورت $i.i.d.$ از دو توزیع $\mathcal{D}^+$ و $\mathcal{D}^-$ به الگوریتم A داده شود، سپس A فرضیه $h \in \mathcal{H}$ را خروجی بدهد به طوری که با احتمال حداقل $1-\delta$ داشته باشیم:

$$\Pr_{x\sim\mathcal{D}^+}[h(x)=0] \leqslant \epsilon \ \ and \ \ \Pr_{x\sim\mathcal{D}^-}[h(x)=1] \leqslant \epsilon$$

الف) ثابت کنید اگر $\mathcal{H}$ با استفاده از مدل پایه‌ی یک توزیعه‌ی PAC قابل یادگیری باشد، با استفاده از مدل دو توزیعه‌ی PAC نیز قابل یادگیری خواهد بود.

ب) (امتیازی) فرض کنید h_0 تابعی است که همیشه 0 خروجی می‌دهد و h_1 تابعی است که همیشه 1 خروجی می‌دهد. ثابت کنید اگر یک کلاس فرضیه $\mathcal{H}$ با استفاده از مدل دوتوزیعه‌ی PAC قابل یادگیری باشد، کلاس فرضیه $\mathcal{H} \cup \{h_0, h_1\}$ نیز بر روی مدل یک توزیعه‌ی PAC قابل یادگیری است.

راهنمایی: می‌توانید از Chernoff bound استفاده کنید.

موفق باشید