

۳.۷ ۳.۶ ۳.۵ ۳.۴ ۳.۳ ۳.۲

۳.۲ می خواهیم بداند که آیا می توانیم با استفاده از ERM برای یادگیری تابع H قابل

افت

استفاده باشد؟ آلودگی به این صورت می گیریم که اگر نمونه مثبت در مجموعه آموزشی

استفاده کنیم از آلودگی بالا

$x = z$	۱
$x \neq z$	۰

که $h(z)$ همان نمونه مثبت است و اگر

هیچ نمونه مثبتی در z نبود، h^- را روی z اعمال می کنیم

(ب) H یک Pac است اگر به ازای $\epsilon \in (0, 1)$ و $D \sim X$ باشد، اگر f

$learnable$

(True label) تمام لیبیل ها را صحتی بپردازند آنگاه آلودگی استفاده شده در مسئله به روش h^-

را برمی بپردازند. اما اگر f به ازای تمام x ها، صحتی نباشد می توان فرض کرد که x صحتی وجود

دارد که لیبیل مثبت برمی بپردازد و اگر x در z باشد آلودگی ما تابع h را برمی بپردازد

حال طبق تعریف Pac می‌گوییم، احتمال حضور h مثبت در K از آن کمتر باشد

(یعنی شانس کمی دارد که h مثبت ظاهر شود و می‌توان گفت Pac است و حال خطای حقیقی کم)

برای اثبات فرض را نقض می‌کنیم یعنی: اگر احتمال حضور h مثبت در K از ϵ

بزرگتر باشد / یعنی احتمالش بالاست (وکی در واقعیت h مثبت در K ظاهر نشود)
 خطای حقیقی بالا

فرض داریم: $P_{S|x \sim D} (L_D(h^+) > \epsilon) = P_{S|x \sim D} (L_D(h) < \delta) < \delta$

$$P_{S|x \sim D} (L_D(h^+) > \epsilon) < (1 - \epsilon)^m \leq e^{-m\epsilon}$$

سخت‌تر شود

$$\Rightarrow e^{-m\epsilon} < \delta \Rightarrow m_H(\epsilon, \delta) \leq \left\lceil \frac{\ln(1/\delta)}{\epsilon} \right\rceil$$

یعنی با sample complexity $m_H(\epsilon, \delta)$ می‌توانیم گرفت که $h \in H$ Pac است
 learnable

3.3 برای اثبات اینکه H و Pac یادگیر بیهوشی را نمی‌گیرند که $X = R^r$ learnable

ما ϵ, δ و γ هفت برای ϵ, δ, γ و یک مجموعه آموزشی S و $m_H(\epsilon, \delta, \gamma) \geq m$ که

استاد از دستور A یک تابع h را برگزیند که $r \leq \|h\|_\infty \leq \frac{1}{\epsilon}$ و h و r تابع

کوچکترین دایره محاطه S است. حال یادگیر به فرم h و r که r و h که

مسئله شده ϵ و δ را شعاع دایره در نظری توپ (شعاع دایره در واقع کمترین شعاع است).

حال یک r' طوری فرض کنید که $r' \leq r$ باشد پس $P(\|h\|_\infty \leq r' \mid \|h\|_\infty \leq r) \geq \epsilon$

حال برای اثبات: $\Rightarrow \delta \leq \epsilon$

$$\delta \leq e^{-\epsilon m} \leq (1-\epsilon)^m$$

$$\Rightarrow m_H(\epsilon, \delta) \leq \left\lceil \frac{\ln(1/\delta)}{\epsilon} \right\rceil \quad \left(\begin{array}{l} \text{پس } H \text{ و } Pac \text{ یادگیر بیهوشی را نمی‌گیرند} \\ \text{و } r=r' \text{ است} \end{array} \right)$$

۳.۵) یعنی X_i ها مستقل هستند و یک هم توزیع هستند: $X_i \sim P_i$

H یک مجموعه فرضیات متناهی از binary classification است. فرضیات f و \bar{f} .

یک سری آموزش S و آنگاه A که $A(S) = h$ و A یک شروع و $\bar{D}_m = \frac{P_1 + \dots + P_m}{m}$

$$P(\exists h \in H \text{ s.t. } \sum_{x \sim \bar{D}_m} |h(x) - f(x)| > \epsilon) < \delta$$

اثبات: فرض با خطای ϵ و δ

$$\sum_{x \sim \bar{D}_m} P[h(x) \neq f(x)] + \dots + P_{x_m \sim P} [h(x) \neq f(x)] > \epsilon$$

$$\Rightarrow \sum_{x \sim \bar{D}_m} P_{x_1 \sim P} [h(x) = f(x)] + \dots + P_{x_m \sim P} [h(x) = f(x)] < 1 - \epsilon$$

✓ حال چون کار با خطای حقیقی است و چیل realizability وجود دارد پس با استفاده از خطای تجربی داریم:

$$P_{S \sim P_1} [L_S(h) = 0] = \prod_{x \sim P_1} P [d(x) = f(n)]$$

$$= \left(\prod_{x \sim P_1} P [h(x) = f(n)] \right)^{1/m}$$

احتمال میانگین برای m مثال

$$\left(\sum_{i=1}^m P_{x \sim P_1} [h(x) = f(n)] \right)^{1/m}$$

$$\Rightarrow P_{x \sim P_1} [h(x) = f(n)] \leq (1 - \epsilon)^m \leq e^{-\epsilon}$$

$$\Rightarrow \frac{1}{m} \left[\sum_{i=1}^m P_{x \sim P_1} [h(x) = f(n)] \right] \leq (1 - \epsilon)^m \leq e^{-m\epsilon}$$

\Rightarrow در اینجا h به روشی توزیع برای آن اعضای H صورت می‌گیرد.

$$P(\exists h \in H, L_P(h) > \epsilon, L_S(h) = 0) \leq |H| e^{-\epsilon m}$$

۲.۴ با فرض Realizability معیار $\min_{h^* \in H} L_D(h^*) = 0$ می‌توانیم

و $m_H(\epsilon, \delta) = \text{sample complexity}$ برای A را می‌توانیم

داده m نمونه تصادفی است اما $L \neq \text{true label}$ است

$$L_D(h) \leq \min_{h^* \in H} L_D(h^*) + \epsilon$$

$$\leq 0 + \epsilon$$

$$\leq \epsilon$$

که نتیجه یادگیری agnostic PAC learning

است و از PAC learning (یادگیری PAC) می‌گویند.

فصل پنجم

الف) ۶- میزان مدل است که خطای حقیقی کمتری داشته باشد.

و فرض می‌کنیم که مجموعه مضارب یکبار برای BMI و BP است پس H_2 خطا

یکبار برای کل جبروت ها که H_5 شود و $H_2 \subseteq H_5$.

حل می‌رسانیم یا بیشتر متغیرهای H_1 ، خطای تقریبی (ϵ_{approx}) کاهش و صفا پیدا

خطای تخمین افزایش می‌یابد و چون توانایی محاسبه دقیق وجود ندارد پس بنابر

فکره تیغ اگام مدل کوچتر خلاصه را انتخاب می‌کنیم $\leftarrow H_2$

و این نکته حائز اهمیت است که زمانی که $Training$ کوچک باشد بهتر است از تیغ اگام set

۷- H_2 استفاده کنیم.

✓ طبق مکالمات الز تعداد Training set ها تا ۱۰۰۰۰ feature ها

باستد آنگاه تعداد نمونه می توان گفت که کافیه و می توان مقدار

دقیق را حساب کرد که از ۱۵ یا ۲۰ استفاده کنیم. (در غیر اینصورت

برای Training set کمتر بهتر است از ۲۰ استفاده کنیم.