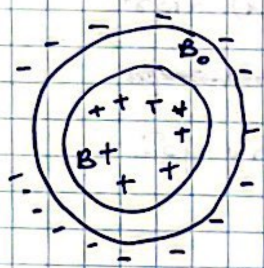
למידה חישובית - תרגיל 2

(1) בהינתן מדגם בגודל n בלתי תלוי נמצא את העיגול הכי קטן שעקבי עם המדגם ונסמן אותו ב B_0 כלומר עבור המדגם $\{a_1,\dots,a_n\}$

לכל a_i $\|a_i\|_2 \le R$. נסמן ב B את הכדור ה"מושלם" הקיים כי אנו במקרה realizable

נשים לב שהטעות היחידה שנעשה היא לסמן נקודות חיוביות כשליליות.



הטעות של האלגוריתם תהיה $e_p(h_R) = P(B_0 \setminus B)$

אם $P[B_0] \le e$ אז $e_p(h_R) = P[B_0 \setminus B] \le P[B_0] \le e$

נניח כעת כי $P[B_0] \ge e$

נסתכל על מרכז B ונסמן את T כך ש $P(T)=\varepsilon$

T מתחילה מקצוות B_0 ומתקדמת פנימה לכיוון המרכז.

אם נניח כי קיימת דגימה a_i כך ש $a_i \in T$ אז נקבל כי $B_0 \setminus B_R \subseteq T$

ולכן $e_p(h_R) = P(B_0 \setminus B_R) \le P(T) = \varepsilon$

באופן שווה אם $e_p(h_R) > \varepsilon$ נובע שכל n הדגימות לא נמצאות ב T

הסיכוי שדגימה לא ב T היא $1-\varepsilon$ ולכן:

$$P(e_p(h_R) > \varepsilon) \le P(\forall i : a_i \notin T) \le \prod_{i=1}^{n} P(a_i \notin T) = (1-\varepsilon)^n$$

נקבל ש $(1-\varepsilon)^n \le e^{-n\varepsilon} \le \delta \Rightarrow n \ge \frac{1}{\varepsilon} \cdot \ln\frac{1}{\delta}$

לכן נסמן $N(\varepsilon,\delta) = \frac{1}{\varepsilon}\ln\frac{1}{\delta}$

עבור כל ε,δ וכל realizable התפלגות P מעל R^d עם labeling $B_0 \in H_{ball}$

כאשר מריצים אלגוריתם שתואר למעלה עם $n \ge N(\varepsilon,\delta)$ דגימות בצורה בת"ל מ P נקבל היפותזה h_R מתוך H_{Ball}

כך ש $P[e_p(h) > \varepsilon] \le (1-\varepsilon)^n \le \delta$. ולכן H_{Ball} היא למידה PAC.

נשים לב כי סיבוכיות המדגם לא תלויה במימד d.

∎

המשך

② בכיוון הראשון נניח כי H היא למידה PAC בתוחלת.

יודעים כי $e_p(A(S)) \geq 0$, ולכן לפי אי-שוויון מרקוב מקבלים כי

$P(e_p(A(S)) > \varepsilon) \le \frac{E[e_p(A(S))]}{\varepsilon}$.

עבור $a\in(0,1)$ יהי δ, ε עבורם $a=\varepsilon\cdot\delta$

. $N(a) = N(\varepsilon\cdot\delta)$

ולכן מתקיים:

$$P(e_p(A(S))>\varepsilon) \le \frac{E[e_p(A(S))]}{\varepsilon} \le \frac{a}{\varepsilon} = \frac{\varepsilon\cdot\delta}{\varepsilon} = \delta$$

כלומר לכל $n > N(a)$ זה מתקיים ולכן H היא למידה PAC.

בכיוון השני נניח כי H למידה PAC. $a\in(0,1)$ לכל $n>N(\varepsilon,\delta)$: (כך ש- $a=\varepsilon+\delta$)

$$E[e_p(A(S))] \overset{\text{לפי הגדרת תוחלת}}{=} E[\overbrace{e_p(A(S)) \mid e_p(A(S))>\varepsilon}^{\le 1}]\, P(\overbrace{e_p(A(S))>\varepsilon}^{\le\delta}) +$$
$$+ E[\overbrace{e_p(A(S)) \mid e_p(A(S))<\varepsilon}^{\le\varepsilon}]\, P(\overbrace{e_p(A(S))<\varepsilon}^{\le 1}) =$$
$$\overset{\text{לפי PAC}}{\le} \delta+\varepsilon = a$$

ולכן עבור $a=\varepsilon+\delta$ H למידה PAC בתוחלת עם אלגוריתם A.

למידה חישובית - המשך

3. נוכיח כי $VC\text{-}dimension(H_k) = 2k$

נראה כי $VC(H_k) \ge 2k$:

נבחר קבוצה בגודל 2k: $x_i = \frac{i}{2k}$ ויהי $s_1, \dots, s_{2k} \in \{0,1\}$ תיוג כלשהו.

עבור כל רצף של $x_i, x_{i+1}, \dots, x_j$ עבורם $y_i = \dots = y_j = 1$ נגדיר אינטרוול $[x_i, x_j]$

ולכן לכל $i \le m \le j$: $h(x_m) = 1$ ולכל נקודה שלא שובצה לרצף תקבל $h(x_m) = 0$.

נוכיח שיש לכל היותר k רצפים בכל 2k נקודות. נניח בשלילה כי יש k+1.

כל 2 רצפים מופרדים לפחות על ידי 0 אחד. [ ] 0 [ ] 0 [ ]

ולכן במקרה הטוב שבו כל רצף מכיל רק איבר אחד נקבל שהרצפים מכילים k+1 איברים

ובנוסף יש k אפסים. סה"כ לפחות 2k+1 נקודות בסתירה לכך שהקבוצה בגודל 2k.

נראה כי $VC(H_k) < 2k+1$:

יהי קבוצה בגודל 2k+1 $x_1, \dots, x_{2k+1}$ נניח בה"כ $x_1 < x_2 < \dots < x_{2k+1}$.

ניתן לסדר את כל איבר 1 ו-0. כלומר $x_i = (-1)^{i+1}$ $(1,0,1,0,\dots,1)$

נניח והצלחנו למצוא k אינטרוולים. אז כיוון שיש k+1 אחדים מעקרון היונים יש

2 אחדים בתוך אינטרוול נסמן אותם x_i, x_j וכל מהגדרת הקבוצה קיים m עבורו

$x_i < x_m < x_j$ ולכן האינטרוול $[x_i, x_j]$ מכיל נקודה שהלייבל שלה 0 בסתירה.

סה"כ קיבלנו כי $VC(H_k) = 2k$

למידה חישובית

(4) נוכיח כי $VC(H_{poly}) = \infty$

נראה כי לכל n קיים קבוצה C של n משתנים שמנותצת ע"י H_{poly}.

נבחר $C = \{(x_1,0), \dots, (x_n,0)\}$ כאשר ה x_i בסדר עולה.

לפי הרמז לכל $x_1,\dots,x_n \in \mathbb{R}$ שונים קיימים $z_1,\dots,z_n \in \mathbb{R}$ שונים

קיים פולינום p מדרגה $n-1$ כך שלכל i מתקיים $p(x_i) = z_i$

יהי $\{s_1,\dots,s_n\} \in \{0,1\}^n$ דיכוטומיה.

לכל i מתקיים אם $s_i=0$ נבחר $(x_i,-1)$ אחרת $(s_i=1)$ $(x_i,0)$

יהי הפולינום p המקיים $p(x_i) = \begin{cases} 0 & s_i=1 \\ -1 & s_i=0 \end{cases}$ (לפי הרמז)

ולכן לכל i

$$h_p(x_i,0) = \begin{cases} 1 & p(x_i)\geq 0 \\ 0 & \text{else} \end{cases} =$$

$$= \begin{cases} 1 & s_i=1 \\ 0 & s_i=0 \end{cases} = s_i \Rightarrow |H_{poly_C}| = 2^{|C|}$$

כלומר קיבלנו כי $VC\dim(H_{poly}) \geq n$ וכיוון ש n לא קבוע וניתן להגדיר כרצוננו אז $VC\dim(H_{poly}) = \infty$

חלק מעשי

הערה: אנו מסתכלים על הקטע $[0,1]$ כלומר אם $x \notin H$ אז $x \in \{(0.2,0.4),(0.6,0.8)\}$

1c) $e_p(h)=E[\Delta(Y,h(x))]$ נחשב תחילה בנפרד: $H=\{(0,0.2),(0.4,0.6),(0.8,1)\}$

$$(*)\quad P(h(x)=1,Y=0)=P(Y=0\mid h(x)=1)\,P(h(x)=1)=\overbrace{P(Y=0\mid x\in H)}^{0.2}\,P(x\in H,h(x)=1)+$$

$$+\overbrace{P(Y=0\mid x\notin H)}^{0.9}\,P(x\notin H,h(x)=1)=0.2\,P(x\in H,h(x)=1)+0.9\,P(x\notin H,h(x)=1)$$

$$(**)\quad P(Y=1,h(x)=0)=P(Y=1\mid h(x)=0)\,P(h(x)=0)=\overbrace{P(Y=1\mid x\in H)}^{0.8}\,P(h(x)=0,x\in H)+$$

$$+\overbrace{P(Y=1,x\notin H)}^{0.1}\,P(x\notin H,h(x)=0)=0.8\,P(h(x)=0,x\in H)+0.1\,P(x\notin H,h(x)=0)$$

אנחנו רוצים להביא למינימום את הביטויים האלו כדי להקטין כמה שיותר את ההסתברות.

אז בעצם נקבל שכדי שהחיתוך בינהם יהי הכי קטן, נרצה הפלט כלומר התוצאה $h(x)=1$ אמ"מ $x\in H$

ואם $x\notin H$ אז $h(x)=0$, כלומר המקטעים הכי טובים הם בדיוק H ו-$k=3$ היא ההיפותזה האופטימלית. והמצב גם עבור H_{10}

<u>נסכם הכל ביחד:</u>

$$e_p(h)=E[\Delta(Y,h(x))]=P(h(x)\neq Y)=(*)+(**)=$$

$$=0.2\,P(x\in H,h(x)=1)+\cancel{0.9\,P(x\notin H,h(x)=1)}+\cancel{0.8\,P(x\in H,h(x)=0)}+0.1\,P(x\notin H,h(x)=0)=$$

המקרים שהתאפסו לא יכולים לקרות עבור ההיפותזה הכי טובה כפי שהראנו למעלה

$$=0.2\cdot P(x\in H)+0.1\,P(x\notin H)=0.2\cdot 0.6+0.4\cdot 0.4=\underline{\underline{0.16}}$$