# מבוא למערכות לומדות
# תרגיל 2

26 באפריל 2017

## 1 מסווג Bayes

תהי $\mathcal{D}$ התפלגות על $\mathcal{X} \times \{0,1\}$ ותהי $g : \mathcal{X} \to \{0,1\}$. נניח ש־$f_{\mathcal{D}}$ כמתואר בשאלה, כלומר $f_{\mathcal{D}}(x) = \begin{cases} 1 & \text{if } \mathbb{P}[y=1|x] \geq 1/2 \\ 0 & \text{otherwise} \end{cases}$ לכל $x \in \mathcal{X}$. צריך להראות ש:

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

מתקיים:

$$\begin{aligned} L_{\mathcal{D}}(h) &= \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[l^{0-1}(h,(x,y))\right] = \mathbb{E}_{(x,y)\sim\mathcal{D}}\left[1_{h(x)\neq y}\right] \\ &= \mathbb{P}_{(x,y)\sim\mathcal{D}}\left[h(x) \neq y\right] = \mathcal{D}\left(\{(x,y) \in \mathcal{X} \times \{0,1\} \mid h(x) \neq y\}\right) \end{aligned}$$

מחסם האיחוד:

$$\mathbb{P}[h(x) \neq y] \leq \sum_{\substack{(x,y)\in\mathcal{X}\times\{0,1\} \\ h(x)\neq y}} \mathcal{D}(x,y)$$

לכן:

$$\begin{aligned} L_{\mathcal{D}}(f_{\mathcal{D}}) &= \mathbb{P}[f_{\mathcal{D}}(x) \neq y] = \mathbb{P}[f_{\mathcal{D}}(x) \neq y | y=0]\,\mathbb{P}[y=0] + \mathbb{P}[f_{\mathcal{D}}(x)=0|y=1]\,\mathbb{P}[y=1] \\ &= \mathbb{P}[f_{\mathcal{D}}(x)=1|y=0]\,\mathbb{P}[y=0] \end{aligned}$$

$$\mathbb{P}[y=1|x] = \mathbb{P}[y=1]$$

## 2 VC Dimension

### 2.1 פונקציית הזוגיות

יהי $n \in \mathbb{N}$. נטען ש־$\mathrm{VCdim}(\mathcal{H}) = n$ כאשר $\mathcal{H} = \{h_I : I \subset [n]\}$. נשים לב: $|\mathcal{H}| = |P([n])| = 2^n$. תחילה נראה שקיימת קבוצה $C$ מגודל $n$ אותה $\mathcal{H}$ מנתצת. נגדיר: $C = \{e_1, \ldots, e_n\}$ כאשר $e_i$ הוא וקטור שכולו 0ים למעט בכניסה ה־$i$, בה ערכו 1. יהי $x \in \{0,1\}^n$. נראה ש־$x \in \mathcal{H}_C$. נגדיר $I = \{i \in [n] : x_i = 1\}$ ונטען ש־$x = (h_I(e_1), \ldots, h_I(e_n))$. אכן:

$$h_I(e_i) = \left(\sum_{j\in I} (e_i)_j\right) \mod 2 = \left(\sum_{j\in I} \delta_{ij}\right) \mod 2 = \begin{cases} 1, & i \in I \\ 0, & i \notin I \end{cases} = \begin{cases} 1, & x_i = 1 \\ 0, & x_i = 0 \end{cases}$$

לכל $i$, ולכן $x = (h_I(e_1), \ldots, h_I(e_n)) \in \mathcal{H}_C$. מכאן ש־$\{0,1\}^n \subset \mathcal{H}_C$ כלומר $|\mathcal{H}_C| \geq 2^{|C|} = 2^n$. ברור ש־$|\mathcal{H}_C| \leq 2^{|C|}$ ולכן $|\mathcal{H}_C| = 2^{|C|}$ כלומר $\mathcal{H}$ מנתצת את $C$. מכאן: $\mathrm{VCdim}(\mathcal{H}) \geq n = \log_2 |\mathcal{H}|$ אבל ראינו בתרגול 3 (תרגיל 2) ש־$\mathrm{VCdim}(\mathcal{H}) \leq \log_2 |\mathcal{H}|$ לכל קבוצת היפותזות סופית, לכן $\mathrm{VCdim}(\mathcal{H}) = n$ כנדרש.

### 2.2 איחוד סופי של קטעים

יהי $k \in \mathbb{N}$. נטען ש־$\mathrm{VCdim}(\mathcal{H}_{k-intervals}) = 2k$. נגדיר $C = \{1, 2, \ldots, 2k\}$. נראה ש־$\mathcal{H}_{k-intervals}$ מנתצת את $C$. יהי $x \in \{0,1\}^{2k}$. נגדיר סדרת קטעים $\{[a_i, b_i]\}_{i=1}^k$ כך שההיפותזה $h$ המוגדרת על ידי איחודם מקיימת $x = (h(1), \ldots, h(2k))$. אם $x = 0^{2k}$ אז ניקח סדרת קטעים כלשהי על המספרים השליליים - ולכן $h(i) = 0$ לכל $i \in [2k]$ ולכן $x \in \mathcal{H}_C$. נניח עתה ש־$x \neq 0^{2k}$ ונגדיר:

$$a_1 = \min\{i \in [2k] : x_i = 1\} - \frac{1}{2}$$
$$b_1 = \min\{i \in [2k] : i > a_1 \text{ and } x_i = 0\} - \frac{1}{2}$$

ולכל $1 < j \leq k$ נגדיר[1]:

$$a_j = \min\{i \in [2k] : i > b_{j-1} \text{ and } x_i = 1\} - \frac{1}{2}$$
$$b_j = \min\{i \in [2k] : i > a_j \text{ and } x_i = 0\} - \frac{1}{2}$$

**הסבר:** כל קטע מכסה רצף שלם של 1ים ב־$x$ - קל לראות שישנם לכל היותר $k$ רצפים של 1ים המופרדים ב־0ים (כאשר $k$ מתקבל למשל מ־$x = (1,0,1,0,\ldots,1,0)$ ובמקרה זה הקטע הראשון - כפי שהוגדר לעיל - הוא $\left[\frac{1}{2}, \frac{3}{2}\right]$ כי 1, ורק הוא, מוכל בו; השני הוא $\left[\frac{5}{2}, \frac{7}{2}\right]$ כי 3 מוכל בו וכו'). רצף של 1ים ב־$x$ מתאים לתיוג ב־1 של מספרים עוקבים ב־$C$ ולכן אפשר לקחת קטע אחד שיכסה אותם (ולכן לא צריך $2k$ קטעים עבור וקטור שכולו 1ים).

אם כך $\mathcal{H}_{k-intervals}$ מנתצת את $C$ כי כל תיוג בגודל $2k$ אפשרי על $C$ ב־$\mathcal{H}$.

עתה, עבור קבוצה $C = \{x_1, \ldots, x_{2k+1}\}$, נטען ש־$y = (1,0,1,0,\ldots,1,0,1) \notin \mathcal{H}_C$ $\{0,1\}^{2k+1} \ni$. נניח בה"כ ש־$x_i < x_{i+1}$ לכל $1 \leq i \leq 2k$. נניח בשלילה שקיימת היפותזה $h \in \mathcal{H}_{k-intervals}$ כך ש־$y = (h(x_1), \ldots, h(x_{2k+1}))$ - כלומר קיימת סדרת קטעים כך שלכל $i$ עבורו $y_i = 1$ קיים קטע בסדרה המכיל את $x_i$. נשים לב שב־$y$ יש $k+1$ קורדינטות בהן $y_i = 1$ לכן יש $k+1$ איברים ב־$C$ אותם $h$ מתייגת 1. כמוכן, לכל זוג $1 \leq i < j \leq 2k+1$ כך ש־$y_i = y_j = 1$ קיים $i < l < j$ כך ש־$y_l = 0$, לכן אין קטע המכיל גם את $x_i$ וגם את $x_j$ - כי הוא היה מכיל את $x_l$ ולכן $h(x_l) = 1$ בסתירה לכך ש־$y = (h(x_1), \ldots, h(x_{2k+1}))$. לכן ישנם $k+1$ קטעים המכילים זרים המכילים את $k+1$ הערכים המתויגים 1, בסתירה לכך ש־$h$ מוגדרת על ידי איחוד של $k$ קטעים. כלומר אין קבוצה מגודל $2k+1$ המנותצת על ידי $\mathcal{H}$ ולכן $\mathrm{VCdim}(\mathcal{H}_{k-intervals}) = 2k$.

כעת נניח ש־$k$ אינו חסום, כלומר $\mathcal{H}$ היא קבוצת ההיפותזות המוגדרת על ידי איחוד סופי כלשהו של קטעים ממשיים. תהי $C = \{x_1, \ldots, x_n\} \subset \mathbb{R}$ קבוצה כלשהי. יהי $y \in \{0,1\}^n$. לכל $i \in [n]$ כך ש־$y_i = 1$ נגדיר את הקטע $[x_i, x_i]$ כלומר קטע המכיל רק את $x_i$. נקבל אוסף סופי של קטעים $\{[x_i, x_i] : y_i = 1\}$. נתבונן בהיפותזה $h$ המוגדרת על ידי איחוד הקטעים הנ"ל. נשים לב שלכל $i$ כך ש־$y_i = 0$ לא קיים קטע בסדרה הנ"ל המכיל את $x_i$ ולכן $h(x_i) = 0$. מאידך, לפי הגדרה, לכל $i$ כך ש־$y_i = 1$ קיים קטע המכיל $x_i$ ולכן $h(x_i) = 1$. לכן $y \in \mathcal{H}_C$ $(h(x_1), \ldots, h(x_n)) =$ ולכן $\mathcal{H}$ מנתצת את $C$. קיבלנו ש־$\mathcal{H}$ מנתצת כל קבוצה סופית לכן $\mathrm{VCdim}(\mathcal{H}) = \infty$ במקרה זה.

### 2.3 חצאי מרחב לא הומוגניים

נגדיר $C = \{e_1, \ldots, e_d, 0\}$. יהי $y = (y_0, \ldots, y_d)^T \in \{\pm 1\}^{d+1}$, נבחר $\mathbf{w} = (y_0, \ldots, y_{d-1})^T$ ו־$b = \frac{1}{2} y_d$. מתקיים:

$$h_{\mathbf{w},b}(e_i) = \mathrm{sgn}(\langle \mathbf{w}, e_i \rangle + b) = \mathrm{sgn}(y_{i-1} + b) \overset{|b| \leq \frac{1}{2}}{=} y_{i-1}$$
$$h_{\mathbf{w},b}(0) = \mathrm{sgn}(\langle \mathbf{w}, 0 \rangle + b) = \mathrm{sgn}(b) = y_d$$
$$\Longrightarrow (h_{\mathbf{w},b}(e_1), \ldots, h_{\mathbf{w},b}(e_d), h_{\mathbf{w},b}(0)) = (y_0, \ldots, y_{d-1}, y_d) = y$$

[1] אם עבור $j$ כלשהו הקבוצה עליה מחפשים מינימום ריקה ניקח ערך כלשהו מימין לקטע $[1, 2k]$

כלומר $HS_d$ מנתצת את $C$, לכן $\text{VCdim}\left(HS_d\right) \geq d+1$.

נניח בשלילה שקבוצה $C = \{x_1, \ldots, x_{d+2}\}$ מנותצת על ידי $HS_d$. יהי $y \in \{\pm 1\}^{d+2}$. אזי קיימים $\mathbf{w} \in \mathbb{R}^d$ ו־$b \in \mathbb{R}$ כך ש־$h := h_{\mathbf{w},b}$ מקיימת: $\left(h\left(x_1\right), \ldots, h\left(x_{d+2}\right)\right)^T = y$. נגדיר: $\mathbf{w}' = \left(w_1, \ldots, w_d, b\right)^T \in \mathbb{R}^{d+1}$ ולכל $i \in [d+2]$ נגדיר $\mathbf{x}_i' = \left(x_1, \ldots, x_d, 1\right)^T \in \mathbb{R}^{d+1}$. מתקיים:

$$\left\langle \mathbf{w}', \mathbf{x}_i' \right\rangle = \left(w_1, \ldots, w_d, b\right) \begin{pmatrix} x_1 \\ \vdots \\ x_d \\ 1 \end{pmatrix} = \left\langle \mathbf{w}, \mathbf{x}_i \right\rangle + b$$

$$\implies \operatorname{sgn}\left(\left\langle \mathbf{w}', \mathbf{x}_i' \right\rangle\right) = \operatorname{sgn}\left(\left\langle \mathbf{w}, \mathbf{x}_i \right\rangle + b\right) = h_{\mathbf{w},b}\left(x_i\right) = y_i$$

לכל $i \in [d+2]$. מכאן קיימת היפותזה $h' := h'_{\mathbf{w}'} \in HHS_{d+1}$ (קבוצת חצאי המרחב ההומוגניים ב־$d+1$ מימדים) כך ש־$h'\left(\mathbf{x}_i'\right) = y_i$ לכל $i$. זה נכון לכל תיוג $y$, לכן $HHS_{d+1}$ מנתצת קבוצה בגודל $d+2$, וזו סתירה לכך ש־$\text{VCdim}\left(HHS_{d+1}\right) = d+1$. לכן $\text{VCdim}\left(HS_d\right) \leq d+1$ ולכן מתקיים שוויון.

## 3 למידות PAC

## 4 פונקציות במספר משתנים

## 5 חלק תכנותי