פתרון תרגיל 4

תומר דלומי, 305249641

19 ביוני 2018

חלק תיאורטי

1 הוכחה:

בהנחה שאכן מתקיים $\forall t \in \mathbb{N} \;\; \varepsilon_t \leq \frac{1}{2} - \gamma$, ראינו בתרגול שההוכחה ש

$$L_S\left(h_s\right) = \frac{1}{m}\sum_{i=1}^{m} 1_{\left[h_s(x_i) \neq y_i\right]} \leq e^{-2\gamma^2 T}$$

שקולה להוכחה כי מתקיים

$$Z_T \leq e^{-2\gamma^2 T}$$

כאשר

$$Z_t = \frac{1}{m}\sum_{i=1}^{m} e^{-y_i \cdot f_t(x_i)}$$

הראינו כי

$$(*) \quad Z_T = \frac{Z_T}{Z_0} = \frac{Z_T}{Z_{T-1}} \cdot \frac{Z_{T-1}}{Z_{T-2}} \cdots \frac{Z_1}{Z_0}$$

וכי מתקיים אי השוויון

$$\forall t \in \mathbb{N} \;\; \frac{Z_{t+1}}{Z_t} \leq e^{-2\gamma^2}$$

לכן - בהסתמך על משוואה $(*)$, ניתן לרשום:

$$Z_T = \left(\frac{Z_{t+1}}{Z_t}\right)^T \leq \left(e^{-2\gamma^2}\right)^T = e^{-2\gamma^2 T}$$

כלומר

$$Z_T \leq e^{-2\gamma^2 T}$$

$\blacksquare$

1.1 הוכחה אחרת

רק עכשיו קלטתי שכנראה התכוונתם לכך שנוכיח את שתי הטענות הבאות:

1.1.1 שבהנתן $f_t = \sum\limits_{p \le t} w_p \cdot h_p$ **מתקיים** $D_i^{t+1} = \frac{e^{-y_i f_t(x_i)}}{\sum\limits_{j=1}^{m} e^{-y_j f_t(x_j)}}$ **:**

נוכיח באינדוקציה -

<u>בסיס האינדוקציה:</u> $t = 1, 2$ אשר עבורו:

$$D_i^1 = \frac{1}{m},\ f_1 = w_1 h_1,\ D_i^2 = \frac{\frac{1}{m} e^{-y_i w_1 h_1(x_i)}}{\sum\limits_{j=1}^{m} \frac{1}{m} e^{-y_j w_1 h_1(x_j)}} = \dots = \frac{e^{-y_i f_1(x_i)}}{\sum\limits_{j=1}^{m} e^{-y_j f_1(x_j)}}$$

<u>צעד האינדוקציה:</u>

$$D_i^{t+1} = \frac{D_i^t e^{-y_i w_t h_t(x_i)}}{\sum\limits_{k=1}^{m} D_k^t e^{-y_j w_t h_t(x_j)}} = \frac{e^{-y_i w_t h_t(x_i)} \cdot \frac{e^{-y_i f_{t-1}(x_i)}}{\sum\limits_{j=1}^{m} e^{-y_j f_{t-1}(x_j)}}}{\sum\limits_{k=1}^{m} \frac{e^{-y_k f_{t-1}(x_i)}}{\sum\limits_{j=1}^{m} e^{-y_j f_{t-1}(x_j)}} \cdot e^{-y_j w_t h_t(x_j)}} =$$

$$= \dots = \frac{e^{-y_i (w_t h_t(x_i) + f_{t-1}(x_i))}}{\sum\limits_{k=1}^{m} e^{-y_k (w_t h_t(x_k) + f_{t-1}(x_k))}} = \frac{e^{-y_i \left(w_t h_t(x_i) + \sum\limits_{p \le t-1} w_p \cdot h_p \right)}}{\sum\limits_{k=1}^{m} e^{-y_k \left(w_t h_t(x_k) + \sum\limits_{p \le t-1} w_p \cdot h_p \right)}} = \frac{e^{-y_i f_t(x_i)}}{\sum\limits_{j=1}^{m} e^{-y_j f_t(x_j)}}$$

■

1.1.2 שמתקיים $\frac{Z_{t+1}}{Z_t} = 2\sqrt{\varepsilon_{t+1}\left(1 - \varepsilon_{t+1}\right)}$ **:**

$$\frac{Z_{t+1}}{Z_t} = \frac{\frac{1}{m} \sum\limits_{i=1}^{m} e^{-y_i f_{t+1}(x_i)}}{\frac{1}{m} \sum\limits_{j=1}^{m} e^{-y_j f_t(x_j)}} = \dots = \sum_{i=1}^{m} \left[\frac{e^{-y_i \sum\limits_{k=1}^{t} w_k h_k}}{\sum\limits_{j=1}^{m} e^{-y_i \sum\limits_{k=1}^{t} w_k h_k}} \right] e^{-y_i w_{t+1} h_{t+1}} \underbrace{=}_{1.1.1} \sum_{i=1}^{m} D_i^{t+1} e^{-y_i w_{t+1} h_{t+1}(x_i)} = \dots$$

נזכיר:

$$\varepsilon_t = \sum_{i=1}^{m} D_i^t \cdot \mathbf{1}_{[y_i \ne h_t(x_i)]}\ ,\ w_t = \frac{1}{2} ln\left(\frac{1}{\varepsilon_t} - 1\right)$$

ועל כן:

$$\dots = e^{-w_{t+1}}\left(1 - \varepsilon_{t+1}\right) + e^{w_{t+1}} \cdot \varepsilon_{t+1} = \frac{1}{\sqrt{\left(\frac{1}{\varepsilon_{t+1}} - 1\right)}}\left(1 - \varepsilon_{t+1}\right) + \sqrt{\left(\frac{1}{\varepsilon_{t+1}} - 1\right)} \cdot \varepsilon_{t+1} = \dots$$

$$\dots = 2\sqrt{\varepsilon_{t+1}\left(1 - \varepsilon_{t+1}\right)}$$

■

2 הוכחה:

2.1 עבור $h \in L(DS,T)$ נראה $t \in DT$ שעומקו T כך שמתקיים $t(x) = h(x)$ $\forall x \in \mathbb{R}^d$:

יהא $h \in L(DS,T)$, ונסמן ב h_i^j את ה $Decision\ stump$ עבור הקואורדינטה ה $j \in [d]$ (כלומר - בהתייחס לאחד ה $features$), ועומק העץ $i \in [T]$. נסמן ב w_i את המשקל המתאים לו.

יהא $t \in DT$ שעומקו T, כך שבכל רמת עומק i שלו מתקיים:

$$w_i \cdot h_i^j(x) < 0$$

את ערכי העלים ניתן להסיק בהנתן המסלול אליהם מן השורש באופן הבא:

$$sign\left(\sum_{i=0}^{T} sign\left(w_i h_i^j(x)\right)\right)$$

מכאן ניתן לראות שמתקיים:

$$\forall x \in \mathbb{R}^d \quad t(x) = sign\left(\sum_{i=0}^{T} sign\left(w_i h_i^j(x)\right)\right) = sign\left(\sum_{i=0}^{T} sign\left(w_i \cdot sign\left((x_j - \theta)\, b\right)\right)\right) =$$

$$sign\left(\sum_{i=0}^{T} w_i \cdot sign\left(\left((x_j - \theta)\, b\right)\right)\right) = h(x)$$

■

2.2 נראה שישנו $t \in DT$ כך שמתקיים $\forall T \quad t \notin L(DS,T)$:

יהא $t \in DT$ שעומקו $T-1$, ונניח שרמתו התחתונה ביותר תגדיר סיווג מהצורה:

$$x_m < x_{m+1} \mid m \in [d-1] \;\;,\;\; if\ true\ 1\ else\ \ -1 \Leftrightarrow sign\left(x_{m+1} - x_m\right)$$

ניתן לראות בנקל שהיפותזת הסיווג במקרה כזה בהכרח איננה $\in L(DS,T)$ / $boosted\ decision\ stump$, שכן בהגדרת DS עבור כל $feature$ שנבחר כציר ה $split$ ישנו ערך θ אשר תלוי אך ורק בו, בניגוד למקרה הנוכחי בו $\theta \equiv x_m$, כלומר - תלוי ב $feature$ אחר.

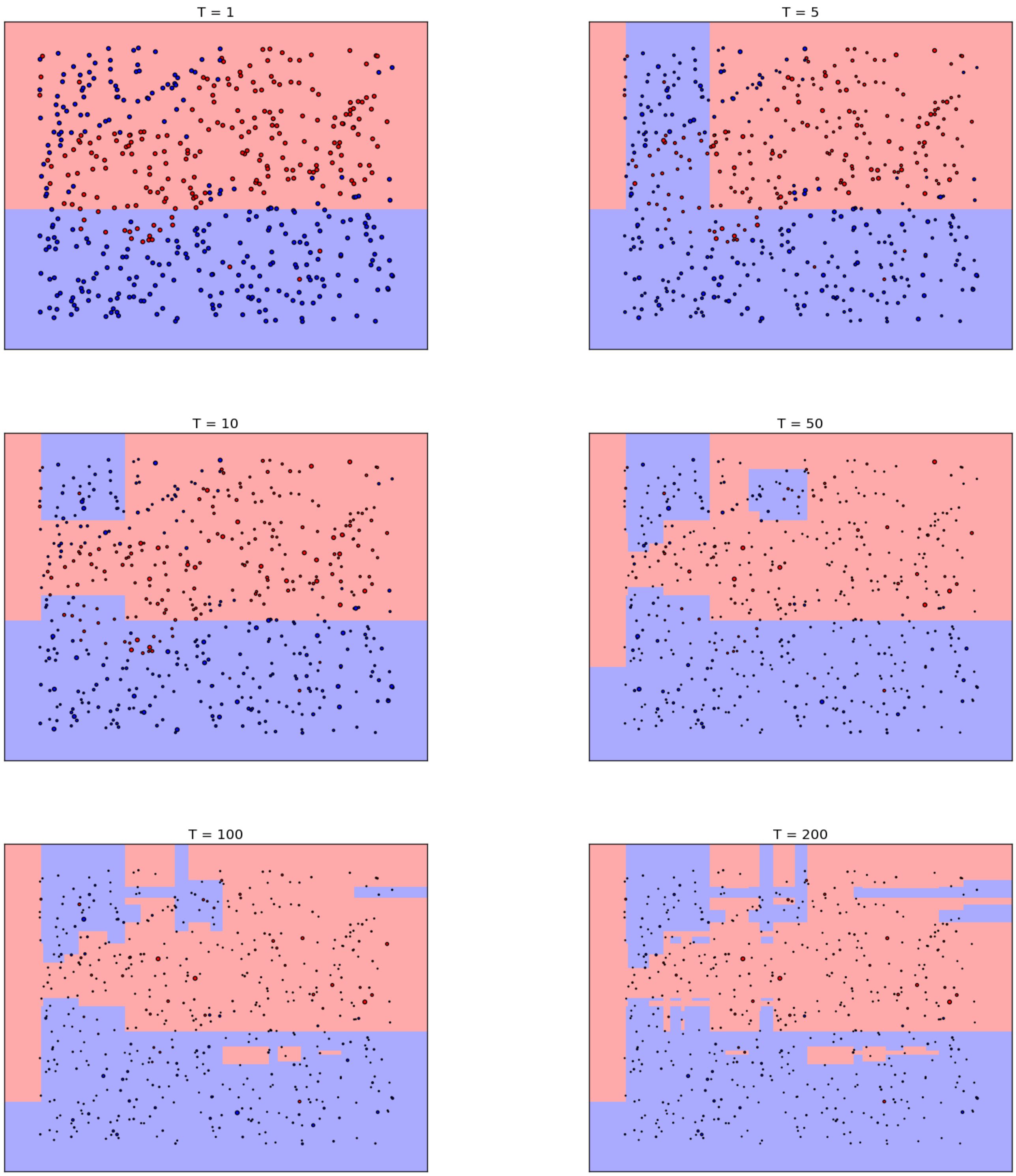
■

חלק מעשי

3 $Adaboost$

3.1 $Training\ error\ and\ validation\ error\ as\ a\ function\ of\ T$

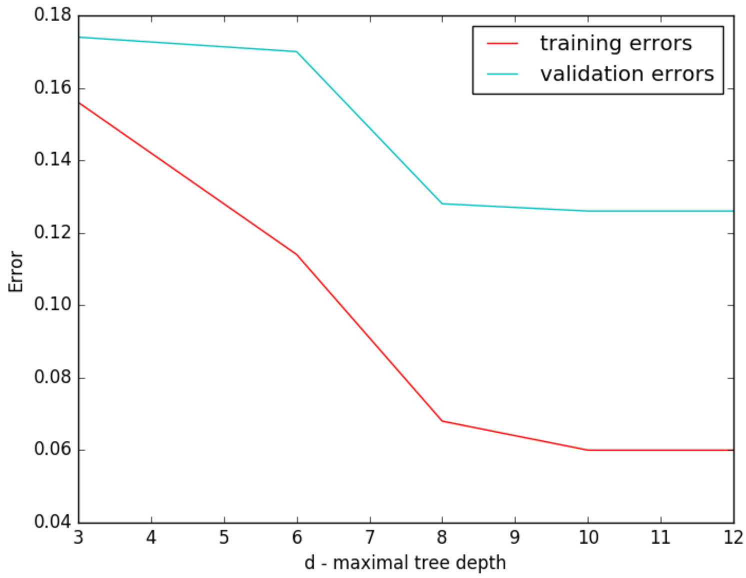
3.2 *Decisions of the learned classifiers with different T values*



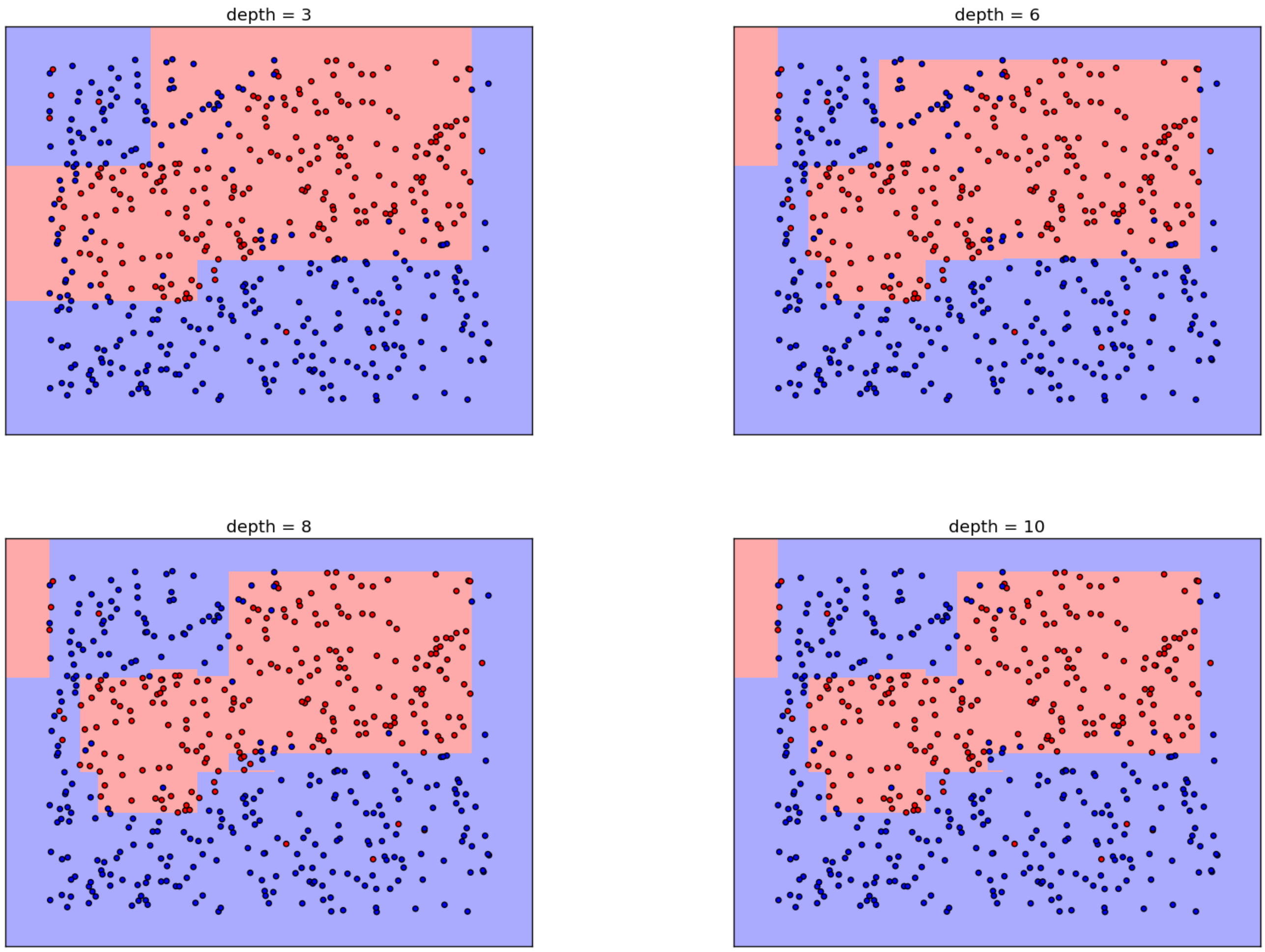
3.3 *minimizing* $\hat{T}$

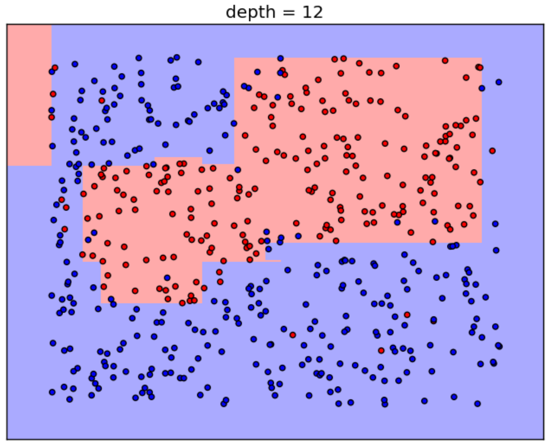
בהרצת הקוד (ראה קוד מצורף) מתקבל הערך $\hat{T} = 50$, ובהתאמה: $test\ error\,[T = 50] = 0.184$

Training error and validation error as a function of depth **4.1**



Decisions of the learned classifiers with different T values **4.2**



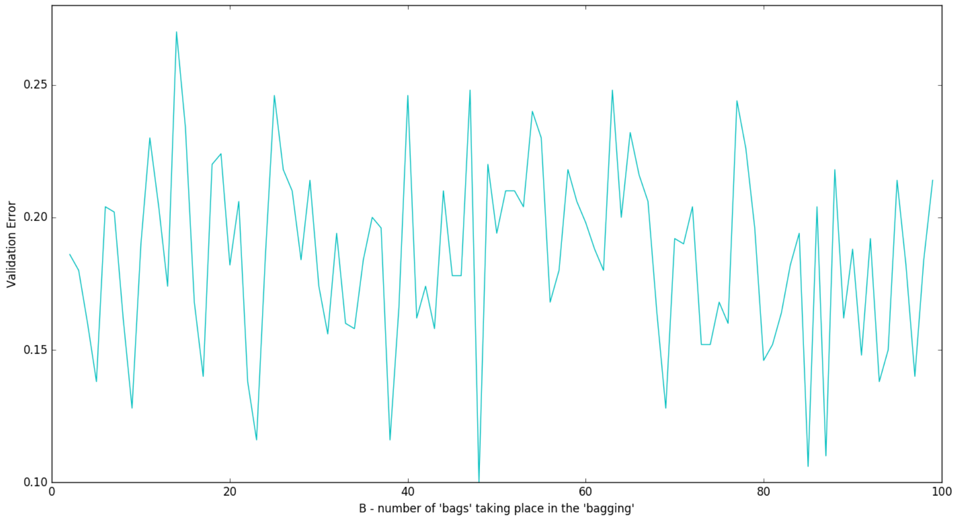


4.3 $minimizing\ d$

בהרצת הקוד (ראה קוד מצורף) מתקבל הערך $depth = 10$, ובהתאמה: $test\ error\,[d = 10] = 0.096$

4.4 $Bonus$

להלן גרף ההתאמה בין שגיאת הולידציה לבין מספר ה"שקים/$bags$" בפתרון הבעיה [1] :



בהרצת הקוד (ראה קוד מצורף) מתקבל הערך $b = 48$, ובהתאמה: $test\ error\,[b = 48] = 0.108$

[1] התוצאה התקבלה עבור עצים שעומקם $d = 10$, על סמך התוצאה שהתקבלה בסעיף 4.3

Spam Dataset and cross validation 5

5.1