

מבוא למכונות לומדות - תרגיל 4

רפאל שקד גרינפלד - 305030868

11 ביוני 2018

1. צ"ל: $\frac{Z_{t+1}}{Z_t} = 2\sqrt{\varepsilon_{t+1}(1-\varepsilon_{t+1})}$. ראשית, נשים לב כי

$$\begin{aligned} D_i^{(t+1)} &= \frac{D_i^{(t)} \exp(-y_i w_t h_t(x_i))}{\sum_{j=1}^m D_j^{(t)} \exp(-y_j w_t h_t(x_j))} = \frac{\frac{1}{m} \prod_{k=1}^t \exp(-y_i w_k h_k(x_i))}{\underbrace{\prod_{k=1}^t \sum_{j=1}^m D_j^{(k)} \exp(-y_j w_k h_k(x_j))}_{\text{we will mark this as K}}} = \\ &= \frac{\exp\left(\sum_{k=1}^t -y_i w_k h_k(x_i)\right)}{m \cdot K} = \frac{\exp\left(-y_i \sum_{k=1}^t w_k h_k(x_i)\right)}{m \cdot K} = \frac{\exp(-y_i f_t(x_i))}{m \cdot K} \end{aligned}$$

ולכן

$$\exp(-y_i f_t(x_i)) = D_i^{(t+1)} \cdot m \cdot K$$

בנוסף,

$$f_{t+1} = \sum_{k=1}^{t+1} w_k h_k = \sum_{k=1}^t w_k h_k + w_{t+1} h_{t+1} = f_t + w_{t+1} h_{t+1}$$

כעת

$$\begin{aligned}
\frac{Z_{t+1}}{Z_t} &= \frac{\frac{1}{m} \sum_{i=1}^m \exp(-y_i f_{t+1}(x_i))}{\frac{1}{m} \sum_{i=1}^m \exp(-y_i f_t(x_i))} = \frac{\sum_{i=1}^m \exp(-y_i (f_t(x_i) + w_{t+1} h_{t+1}(x_i)))}{\sum_{i=1}^m \exp(-y_i f_t(x_i))} = \\
&= \frac{\sum_{i=1}^m D_i^{(t+1)} \cdot m \cdot K \exp(-y_i w_{t+1} h_{t+1}(x_i))}{\sum_{i=1}^m D_i^{(t+1)} \cdot m \cdot K} = \sum_{i=1}^m D_i^{(t+1)} \exp(-y_i w_{t+1} h_{t+1}(x_i)) \\
&= \sum_{i=1}^m D_i^{(t+1)} \exp(-w_{t+1}) 1_{[y_i = h_{t+1}(x_i)]} + \sum_{i=1}^m D_i^{(t+1)} \exp(w_{t+1}) 1_{[y_i \neq h_{t+1}(x_i)]} \\
&= \exp\left(-\frac{1}{2} \log\left(\frac{1}{\varepsilon_{t+1}} - 1\right)\right) \sum_{i=1}^m D_i^{(t+1)} 1_{[y_i = h_{t+1}(x_i)]} + \exp\left(\frac{1}{2} \log\left(\frac{1}{\varepsilon_{t+1}} - 1\right)\right) \sum_{i=1}^m D_i^{(t+1)} 1_{[y_i \neq h_{t+1}(x_i)]} \\
&= \left(\frac{1}{\varepsilon_{t+1}} - 1\right)^{-\frac{1}{2}} \cdot (1 - \varepsilon_{t+1}) + \left(\frac{1}{\varepsilon_{t+1}} - 1\right)^{\frac{1}{2}} \varepsilon_{t+1} \\
&= \sqrt{\frac{\varepsilon_{t+1}}{1 - \varepsilon_{t+1}}} (1 - \varepsilon_{t+1}) + \sqrt{\frac{1 - \varepsilon_{t+1}}{\varepsilon_{t+1}}} \varepsilon_{t+1} = 2\sqrt{\varepsilon_{t+1} (1 - \varepsilon_{t+1})}
\end{aligned}$$

2. (א) תהי $h \in L(DS, T)$, $h = \text{sign}\left(\sum_{t=1}^T w_t h_t(x)\right)$. לכל t , יהי $j_t \in [d]$ נקודת הפיצול של $h_t = (x_{j_t} - \theta_t) \cdot b_t$.

נגדיר עץ $tree(x)$ מאוזן בעומק T כך שברמה t , נקודת הפיצול של העץ תהיה $x_{j_t} \geq \theta_t$. אז נשים לב כי בהינתן

עלה l , לכל $x_1, x_2 \in \mathbb{R}$ שמגיעים ב- $tree$ ל- l , מתקיים לכל t

$$\text{sign}(x_{1j_t} - \theta_t) = \text{sign}(x_{2j_t} - \theta_t)$$

בפרט נובע כי

$$h_t(x_1) = h_t(x_2)$$

ולכן

$$h(x_1) = \text{sign} \sum_{t=1}^T w_t h_t(x_1) = \text{sign} \sum_{t=1}^T w_t h_t(x_2) = h(x_2)$$

לכן נקבע כי $tree(x_1) = h(x_1)$ ונקבל שזה אכן מוגדר היטב.

(ב) נגדיר את העץ הבא: שעבור $X_i \leq 1$ מחזיר 1 אחרת גם 1. (עץ שמדמה היפוטזה נאיבית בצורה של עץ...)

נטען כי אין היפותזה ב- $L(DS, T)$ שמסווגת כמוהו לאף T . יהי T כלשהו ויהי $h \in L(DS, T)$.

נגדיר $sign(0) = 1$ ונניח בשלילה כי $\exists t \in \mathbb{N} :: t \in L(DS, T)$ כעת, נסדר את Θ_i לפי הקורדינטות המכריעות ונראה כי קיים $X \in \mathbb{R} :: t(x) = -1$ והיות והעץ תמיד חיובי נסיק כי קיימת סתירה.

$$\forall x_i \in X$$

נבחר את Θ בקור' Θ_i כך שמשקלה בהיפוטזה DS מקיים שהוא המקסימלי עבור Θ ואזי

$$if(b = 1) : we will define (X_i < \Theta)$$

$$if(b = -1) : we will define (X_i > \Theta)$$

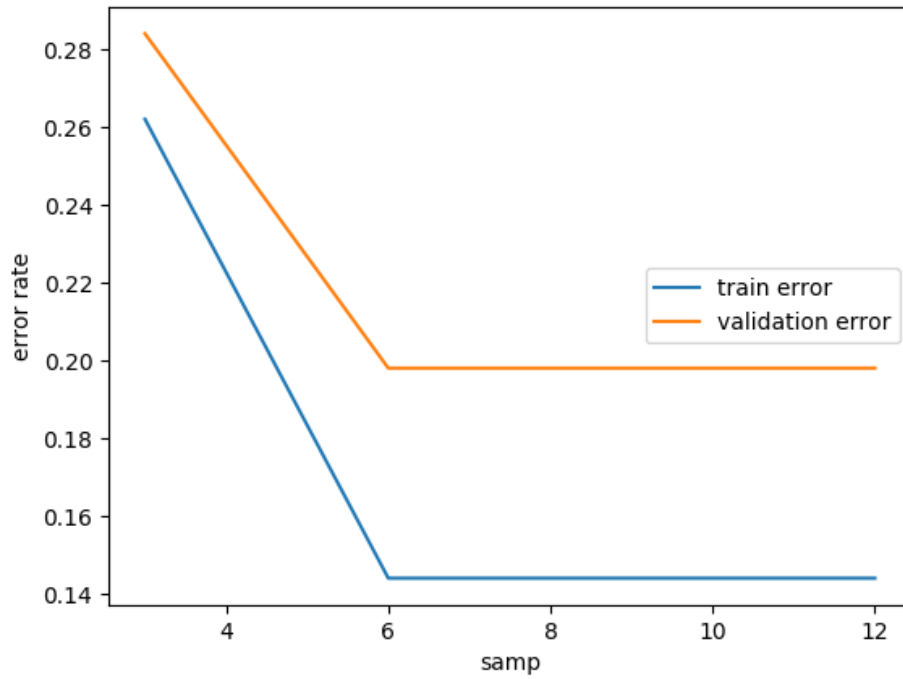
ואזי נקבל שלכל קורדינטה i ב- X סכום ההיפוטזות שמשתמשות בהיפוטזה יהיה בהכרח שלילי ולכן

$$t(x) = sign(\sum_{i=1}^T h_i(X)W_i) = -1$$

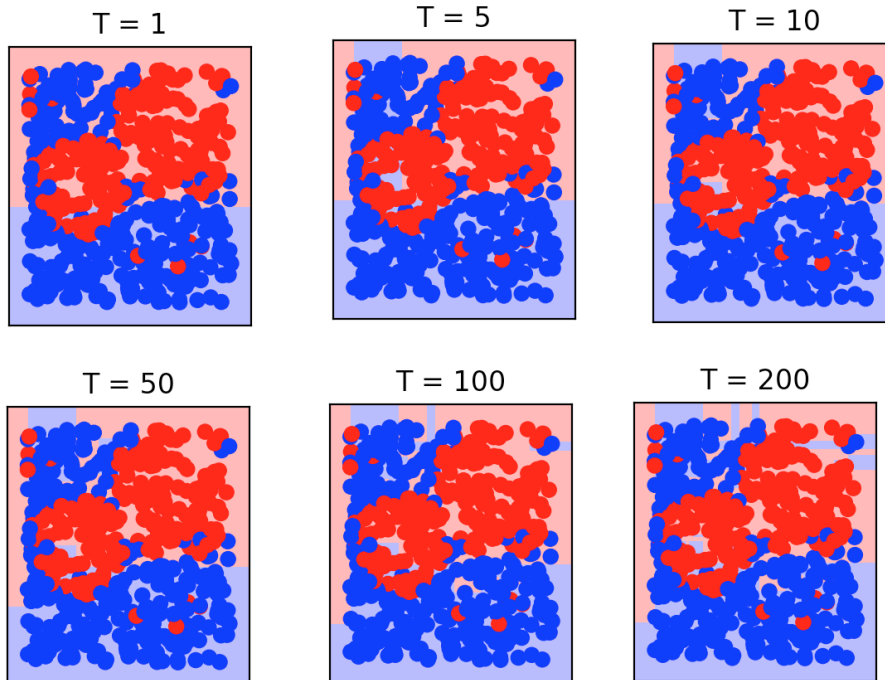
בסתירה.

3.

a. גרף טעות האימון וטעות הולידציה



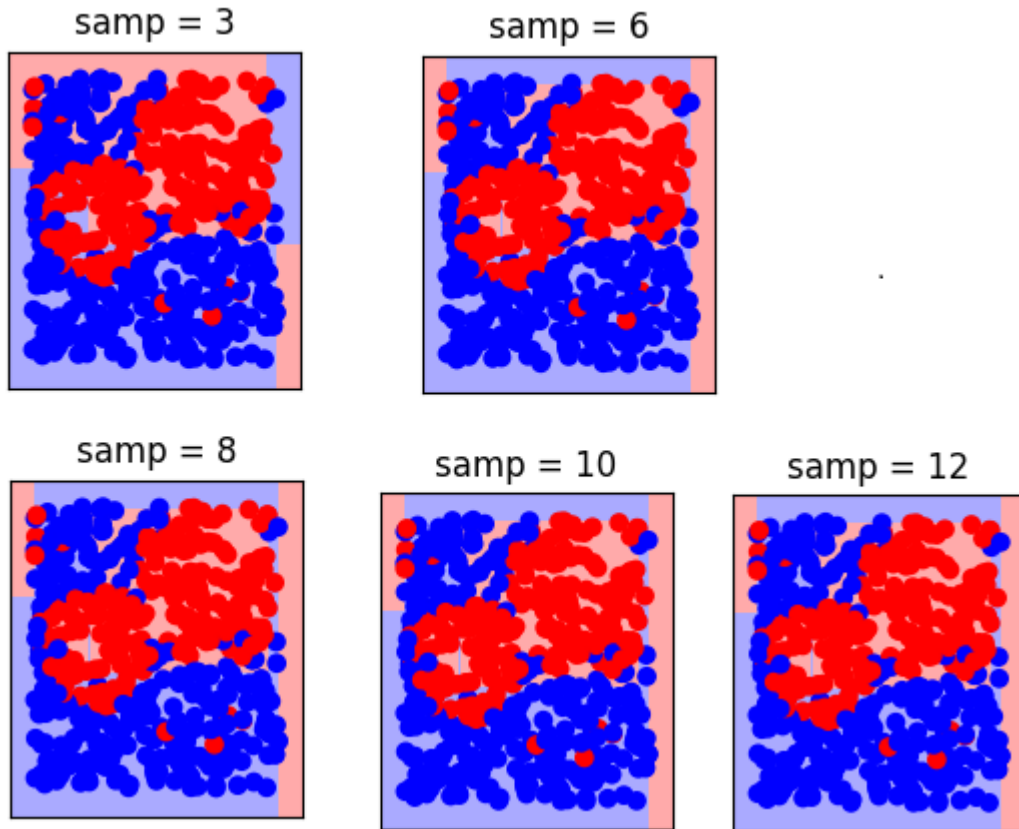
b. גרף ההחלטות עבור מספרי הלומדים T המבוקשים:



c. בכדי למזער את טעות הולידציה, קיבלתי שיש להיות 50 לומדים. כאשר טעות המבחן היא 0.192

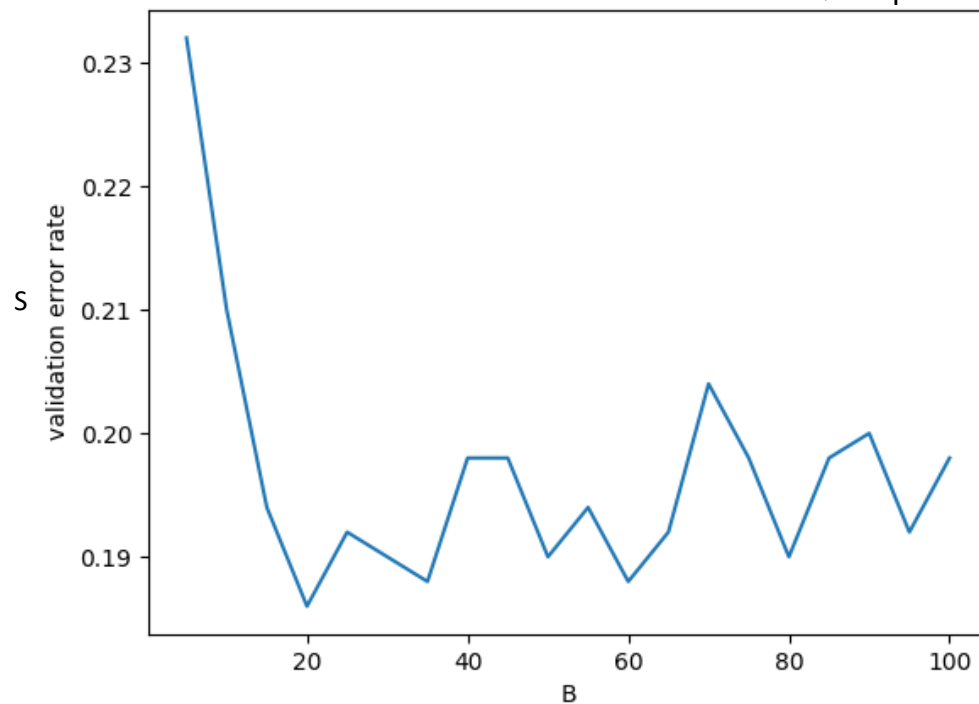
.4

a. טעות הוילדציה הטובה ביותר הייתה כאשר d הוא 6 עם טעות של 0.178
הגרף של הטעויות של האימון והווידוא ביחס לגודל העץ:



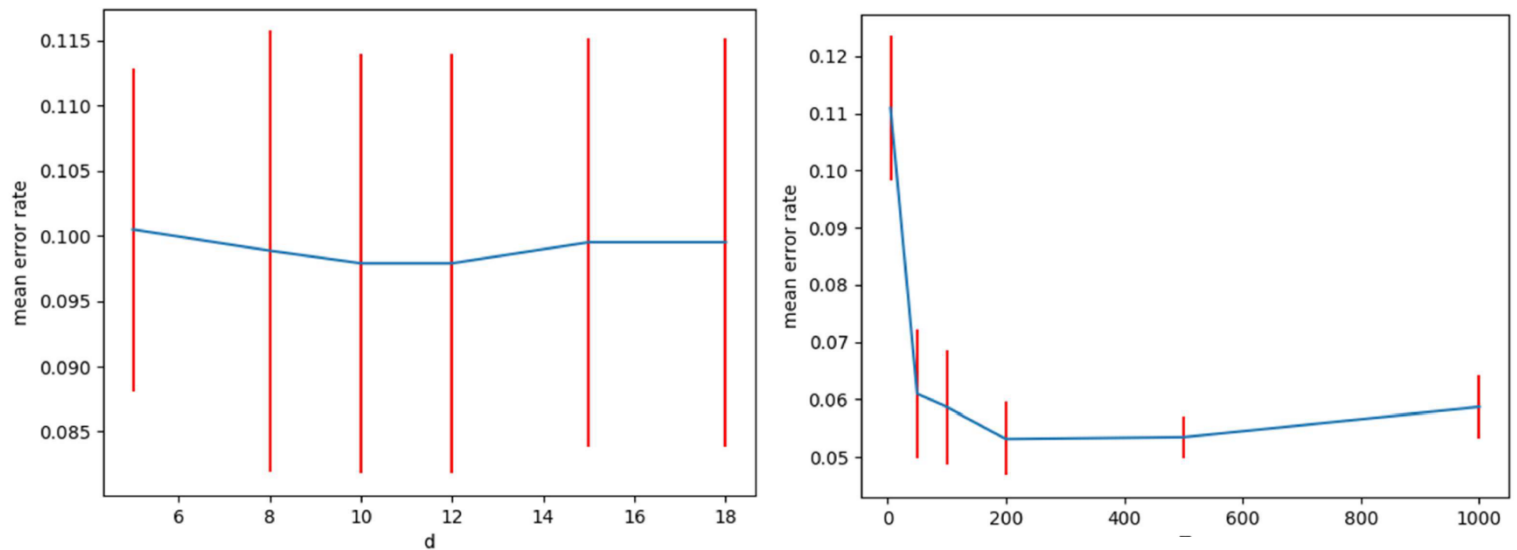
b. Bagging bonus

טעות הוילדציה המינמלית התקבל עם 25 חזרות כאשר הטעות היא 0.17
גרף הטעות:



.5.

a. ניתן לראות בגרף כי טעות הולידציה המינימלית התקבלה בעבור פרטר T 200 כאשר הטעות היא 0.05 ו העומק המקסימלי של העת המספק את הטעות המינמלית הוא 10 עם טעות מדגם של 0.09 הגרפים שנוצרו באמצעות ה 5 fold cross validation לבחירת הT ועומק העץ הטובים ביותר לאלגוריתם adboost ועץ ההחלטה בהתאמה:



b. השונות בטעות הולידציה הינה גבוהה כאשר קיים הבדל קטן יחסית ביחס לעומק עץ. בכדי למנמם את השונות ניתן יהיה להפעיל את אלגוריתם ה – bagging על העומק הנבחר ובכך להוריד את השונות ומכך המסווג יהיה בעל השונות המינמלית.