# פרק 2- האלגוריתם הבסיסי: אדה-בוסט בינארי

בפרק זה נתאר את האלגוריתם הבסיסי המשמש ללימוד מסווגים בינאריים. בעיות אלגוריתמיות, תוצאות התכנסות אלגוריתמית. פרק על מסווגי בסיס ודוגמה פשוטה.

## 2.1 הבסיס

דאטה, טווח שגיאות, מזעור שגיאות.

## 2.2 אלגוריתם אדה-בוסט הבינארי

אדה-בוסט הוא שיטת הרכב (או מטה-למידה) הבונה מסווג באופן איטרטיבי. בכל איטרציה הוא קורא לאלגוריתם למידה פשוט (המכונה הלומד הבסיסי) המחזיר מסווג ומתאים לו מקדם משקלי. הסיווג הסופי ייקבע על ידי "הצבעה" משוקללת של מסווגי הבסיס. ככל שהשגיאה של מסווג הבסיס תהיה קטנה יותר כך יהיה משקלו בהצבעה הסופית גדול יותר. מסווגי הבסיס צריכים להיות רק מעט טובים יותר מניחוש אקראי (ומכאן נובע שמם השני – מסווגים חלשים), מה שמאפשר גמישות רבה בעיצוב קבוצת מסווגי הבסיס.

עבור התיאור הפורמלי של אדה-בוסט, תהי = {(,),…,(,)} קבוצת האימון. האלגוריתם רץ T איטרציות, כאשר T הוא ההיפר-פרמטר היחידי הקבוע מראש באדה-בוסט. הוא יכול להיקבע על ידי, למשל, אימות צולב. בכל איטרציה t = 1,…,T אנו בוחרים מסווג בסיס מהקבוצה H של מסווגים ומתאימים את המקדם שלו . בגרסתו הפשוטה ביותר של האלגוריתם, H היא קבוצה סופית של מסווגים בינאריים מהצורה , והלומד הבסיסי מבצע חיפוש מקיף מעל H בכל איטרציה. הפלט של אדה-בוסט הוא פונקציה מפלה הבנויה כהצבעה משוקללת של מסווגי הבסיס.

)2.1) .

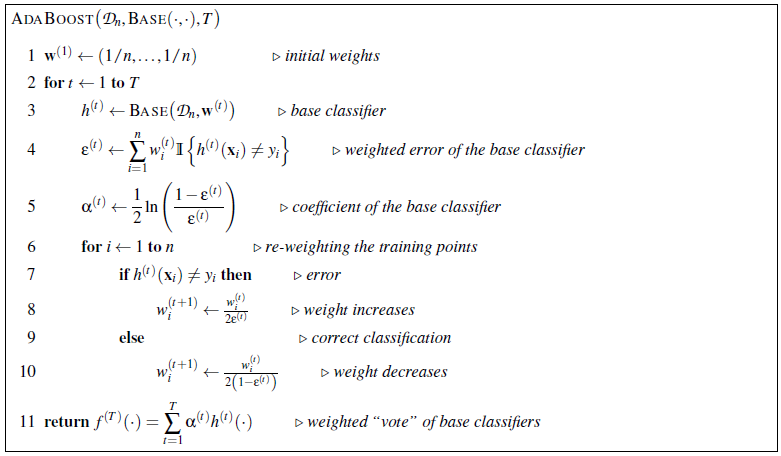
הסימון משמש כסיווג הסופי של x.

האלגוריתם שומר על התפלגות משקל = (,….,) מעל נקודות הדאטה[[1]](#footnote-1). המשקלים מאותחלים באופן אחיד בשורה 1 ומעודכנים בכל איטרציה בשורות 7-10 (תרשים 2.1). מטרתו של הלומד הבסיסי היא למזער את השגיאה הממושקלת.

)2.2) *[[2]](#footnote-2)*

המקדם של נקבע בשורה 5 להיות:

(2.3)



תמונה 2.1: הפסאודוקוד של אדה-בוסט. = {(,),…,(,)} הוא סט האימון, BASE() הוא הלומד הבסיסי, ו-*T* היא מספר האיטרציות.

המשקלים נותרים מנורמלים לאורך כל ריצת האלגוריתם, כלומר לכל t:

)2.4)

על מנת לראות זאת, ראשית יש להבחין כי האתחול מבטיח כי יתקיים .

אז, בהנחה ש , מתקיים כי:

)2.5)

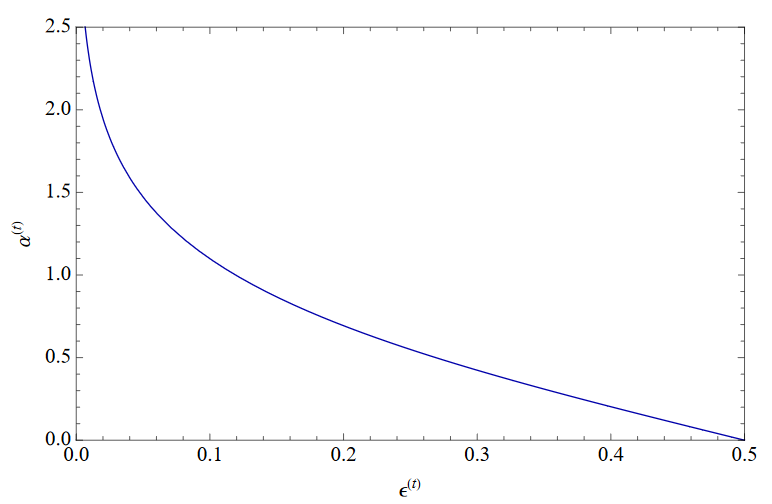
)2.6)

)2.7)

ב- )2.5) הפעלנו את חוקי עדכון המשקלים כפי שמופיעים בשורות 8 ו-10, ב- )2.6) השתמשנו בעובדה כי  
 וב-)2.7) השתמשנו בהגדרה של הנמצאת ב-)2.2) ובהנחה האינדוקטיבית של .

מאחר והמשקלים מנורמלים, הוא מספר ממשי בין 0 ל-1. ובנוסף, אם סגורה לחיסור[[3]](#footnote-3), נוכל להבטיח ש, אחרת, נאלץ להפוך את הסימנים ולהשתמש ב-*-h(t)* במקום ב-*h(t)*.

המסקנה הראשונה מכך שמתקיים היא שמתקיים היא תמיד אי-שלילית. המקדם גדל באופן מונוטוני ככל שערכו של קטן (תמונה 2.2), משמעות הדבר היא כי ככל שמסווג הבסיס *h(t)* טוב יותר, כך משקלו ב"הצבעה" על הסיווג הסופי גדול יותר. ככל ששגיאת הבסיס שואפת ל-0, כך המקדם של שואף לאינסוף. באופן פורמלי, מסווג בסיס חזק עם יקבל, באופן אוטומטי, מקדם אינסופי, כך שישלוט בשילוב הלינארי )2.1). כעקרון, מאחר ומטרת אדה-בוסט היא למזער את שגיאת האימון, משמעה כי מטרת האלגוריתם הושגה, כך שזה נורמלי לסיים את ריצת האלגוריתם. בווריאנטים שונים של אדה-בוסט יהיה זה שימושי להמשיך את ריצת האלגוריתם אף אם .



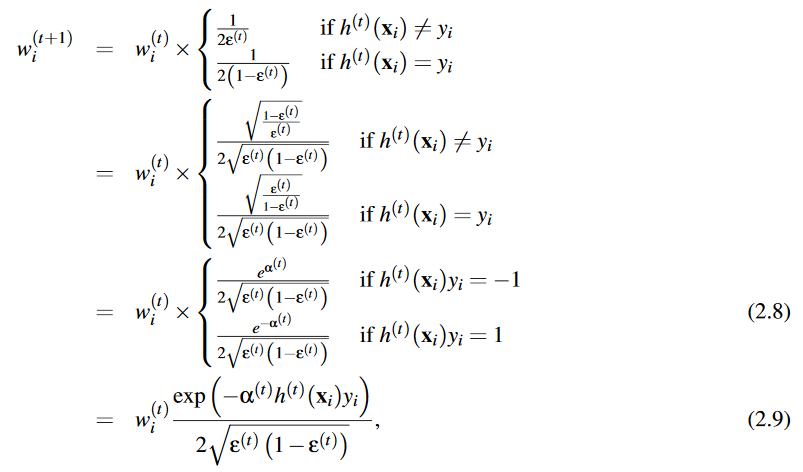
תמונה 2.2: מקדם הבסיס כפונקציה של שגיאת הבסיס . כאשר , , ומתקיים ככל ש- .

המסקנה השנייה מכך שמתקיים היא שכשאר המשקלים *wi* של נקודות האימון מעודכנות בשורות 7 עד 10, המשקלים של נקודות שסווגו באופן שגוי עולים (שורה 8), והמשקלים של נקודות שסווגו נכונה יורדים (שורה 10). ככל שהאלגוריתם מתקדם, המשקלים של נקודות שסווגו לא נכונה באופן תדיר ייטו להיות גדולים, כך שמסווגי הבסיס יתרכזו יותר ויותר בנקודות ה"קשות" הללו.

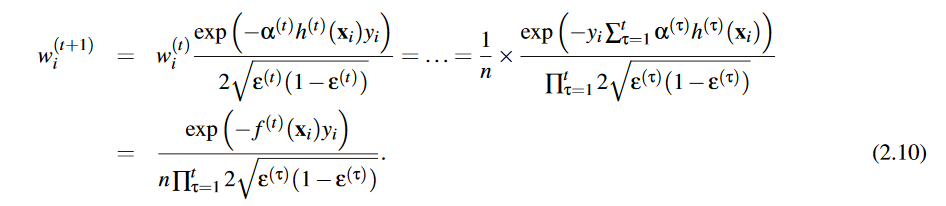
לבסוף, על-ידי בחינת הפיתוח של משוואת הסכום (2.5), נוכל להבחין כי טעות הבסיס של מסווג הבסיס ה-t, *h(t)* הוא , כך שלעולם לא נבחר באותו מסווג בסיס בשתי איטרציות עוקבות (אלא אם אנו עוסקים במקרה מנוון).

## 2.3 אנליזה בסיסית

ראשית אנו נאחד את שני המקרים בנוסחת עדכון המשקולות (שורות 7-10).



כאשר ב- (2.8) אנו משתמשים בהגדרה (2.3) של המקדם ובעובדה שערכיהם של ו- הם {-1,1}.  
על ידי שימוש בנוסחה המאוחדת (2.9), ההגדרה (2.1) של הפונקציה המפלה הסופית של , ואתחול האחיד של המשקלים בשורה 1, אנו יכולים לבטא את המשקלים באופן לא רקורסיבי באופן הבא:

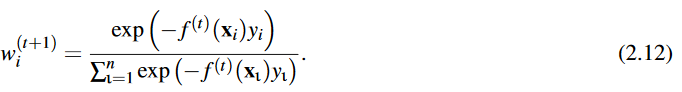
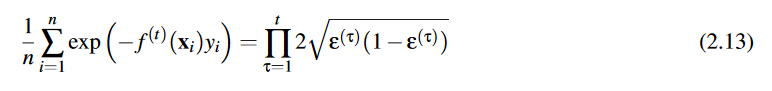


ראשית, הנוסחה הלא רקורסיבית הזו מראה לנו כי המשקליםל פרופורציונליים לביטוי .

המעריך השלילי הוא ההפרש הלא מנורמל, המוגדר באופן כללי עבור הפונקציה המפלה ונקודת (x,y) כך:

(2.11)

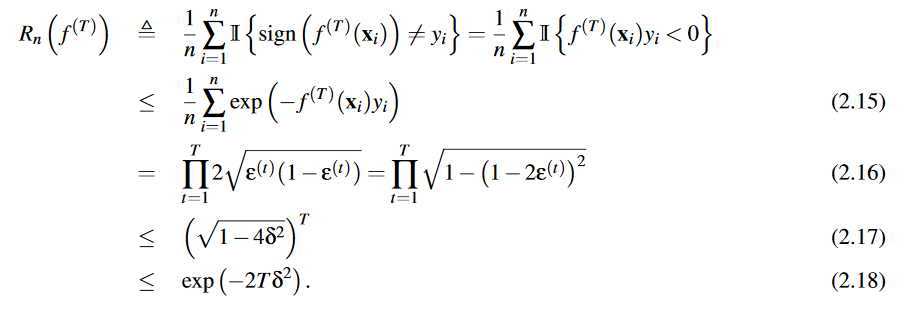
הכמות הזו תשחק תפקיד קריטי בניתוח מבוסס הפרש של הכללת שגיאה.  
לעת עתה, ברור כי:  
1. הסימן של מציין האם מסווגים נכונה על ידי או לא.  
2. ככל ש גדול יותר, כך הסיווג של יותר וודאי, לכן באופן אינטואיטיבי זה הגיוני שהמשקל קטן באופן מונוטוני עם .

שנית, מ- (2.7) אנו יודעים שסכום המשקלים הוא 1. כך שמהנוסחה הלא רקורסיבית (2.10) נובע כי   
  
על ידי השוואת (2.10) ו- (2.12) אנו משיגים באופן אוטומטי את משוואת המפתח   


של טענת ההתכנסות.

טענה 1: אם קיים קבוע חיובי עבורו שגיאת הבסיס חסומה מלמעלה על ידי

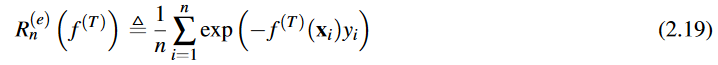
(2.14) .

אז לכל t , שגיאת האימון הופכת ל- 0 לאחר לכל היותר איטרציות.  
הוכחה:  
ראשית חוסמים מלמעלה את שגיאת האימון על ידי:  
  


ב- (2.15) אנו משתמשים ב , (2.16) הוצג ב-(2.13). ב-(2.17) אנו מפעילים את ההנחה כי ו-(2.18) נובע מ- . הטענה הנובעת מאופייה הדיסקרטי של שגיאת האימון: כאשר הגבול העליון יורד מתחת ל, השגיאה חייבת להיות 0.

החשיבות המרכזית של טענה 1 היא שהיא נותנת לנו תשובה חיובית להשערת הגברת הלמידה של PAC: ניתן לבנות לומד חזק (עם סיכון שואף ל-0) על ידי שילוב של מספר קטן של לומדים חלשים (עם סיכונים הקטנים מ- רק במעט). מהירות ההתכנסות המעריכית של שגיאת האימון, הרומזת כי מספר הלומדים החלשים הוא , משחקת תפקיד מכריע בהוכחת ההשערה.

הגבול העליון (2.16) מסביר, למעשה, את מטרת לומד הבסיס והיא למזער את . כמו כן ברור כי על ידי מזעור אנו לא ממזערים באופן ישיר את שגיאת האימון , אלא רק דרך הגבול העליון שלה



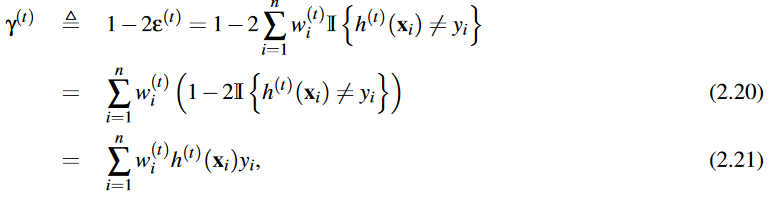
הנקרא הסיכון האמפירי המעריכי. מזעור הגבולות העליונים של עקומה כזאת במקום של שגיאות האימון היא גישה נפוצה בתחום למידת המכונה משתי סיבות:

1. מזעור הוא, לעיתים תכופות, קשה חישובית אפילו עבור פונקציות פשוטות.
2. מזעור הפסדים שוליים הקטנים מונוטונית הוא מוצדק לעיתים תכופות על ידי טיעונים הסתברותיים: אנו משיגים שוליים גדולים יותר ושגיאת הכללה (סיכון) טובה יותר על ידי "החלקת" הסיכון האמפירי.

התחום העליון (2.16) מדגים בנוסף מאפיין חשוב של אדה-בוסט: טבעו החמדני של בסיס האופטימיזציה. אכן, אנו לא מצמצמים גלובלית את הסיכון המעריכי על פני הצירופים הלינאריים של אלא בכל איטרציה *t* אנו מוסיפים את הלומד הבסיסי ואת המקדם שלו שהוא האופטימלי בהנתן המסווג שנבנה עד עתה. פירוש כללי זה של אדה-בוסט כמבצע אופטימיזציה חמדנית של הסיכון המעריכי ייצרה מספר הכללות והרחבות.

התנאי (2.14) הוא למעשה חזק יותר משהוא נראה: אפילו עבור סט-דאטה הוא דורש כי שגיאת הבסיס תהיה פחות מ- לכל התפלגות משקלים *w(t)*. למעשה, זה מאוד אפשרי שלא ניתן יהיה לספק את (2.14): כל שדרוש הוא סט-דאטה שאינו ניתן להפרדה כל ידי אף צירוף לינארי מעל .במקרה כזה שגיאת האימון יכולה בבירור לא להגיע ל-0 ושגיאת הבסיס נוטה לחצי. בגבול, התפלגות המשקלים היא כזו שעבור כל , שגיאת הבסיס היא . מצד שני, אם ניתן להפרדה על ידי צירוף לינארי מעל אז ניתן להוכיח כי (2.14) תמיד ניתנת לסיפוק, וכי הערך האמיתי של (וכך גם מהירות ההתכנסות) מקושר לגבול המפריד בין שתי המחלקות.

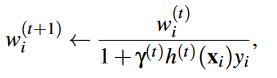
הביטוי אשר מופיע ב-(2.16) נקרא הגבול של מסווג הבסיס , המסומן באמצעות . במובן מסויים, זהו סכום יותר טבעי מאשר השגיאה וייקל את הסימון במספר הצהרות וגזרות. קל לראות כי:



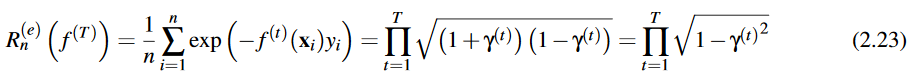
כאשר (2.20) נובע מ- ו-(2.21) הוא אמת היות וערכיהם של ו- הם {-1,1}. הגבול של החלטה אקראית הוא 0 (בממוצע), כך ש- מכמת עד כמה עדיף על פני בחירה אקראית. מספר נוסחאות של האלגוריתם הבסיסי יכולות להיות מפושטות או להיות יותר "סימטריות" בשימוש בגבול במקום בשגיאה. למשל, מקדם הבסיס (2.3) יכול להיכתב מחדש כ:



עדכון המשקל בשורות 7 עד 10 ניתן לפישוט ל:

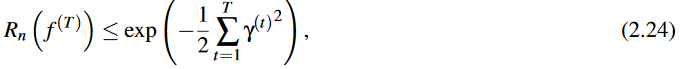


והסיכון המעריכי(2.13) יכול להיות מבוטא כ:



ההגדרה של השוליים (2.11) מספקת פירוש נוסף עבור (2.21): הגבול הוא הממוצע השולי (הממושקל) של מסווג הבסיס . למעשה, הגבול קשור באופן הדוק לשוליים.

כהערה סופית, שימו לב כי הגבול העליון (2.18) ניתן לחידוד על ידי שימוש בגבולות הא-פוסטריוריים במקום גבולותיהם התחתונים הא-פריוריים עבור



כך אדה-בוסט מבטיחה להשיג טעות אימון אפסית כאשר גדול מ-.

1. על מנת למנוע בלבול, מעתה נתייחס למשקלי מסווגי הבסיס כמקדמי α(t) ונשמור את המונח משקל עבור המשקלים wi של נקודות הדאטה. [↑](#footnote-ref-1)
2. הפונקציה המציינת היא 1 אם הארגומנט A עם ערך true ו-0 אחרת. [↑](#footnote-ref-2)
3. הכוונה, אם אז . התכונה הזאת מסופקת על ידי רוב מחלקות המסווגים הפרקטיות, ואנו נניח זאת מכאן והלאה. אם אינה סגורה לחיסור ולא נוכל למצוא מסווג בסיס המקיים , אז האלגוריתם יסיים ויחזיר . במקרה המנוון שבו , נוכל להמשיך בלולאה אך מאחר ו לכל , נצטרך לסיים את ריצת האלגוריתם. [↑](#footnote-ref-3)