האלגוריתם הבסיסי: ADABOOST בינארי.

**2.1 הבסיס**

מידע, שגיאות, שוליים, צמצום בעיות.

**2.2 האלגוריתם ADABOOST הבינארי**

ADABOOST היא מטודת הרכבה (או למידת-מטא) שבונה קבוצת סיווג בצורה איטרטיבית.

בכל איטרציה, היא קוראת לאלגוריתם למידה בסיסי, שמחזירה קבוצת סיווג, ומוסיפה לו משקל מקדם. הסיווג הסופי ייקבע ע"י הצבעה משוקללת של הסיווג הבסיסי.

ככל שהשגיאה של הסיווג הבסיסי קטנה, ככה משקלו גדול יותר לשכלול הסופי.

הסיווג הבסיסי צריך להיות מעט יותר טוב מבחירה רנדומלית לגמרי, מה שמביא גמישות גדולה לעיצוב קבוצת הסיווג הבסיסית.

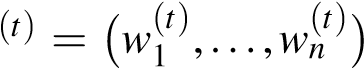
לתיאור הרשמי של ADABOOST, תהי קבוצת האימון-  


האלגוריתם רץ T איטרציות, כאשר T הוא ההיפר-פרמטר היחיד של האלגוריתם ADABOOST אשר מוגדר מראש. בכל איטרציה *t* = 1,*...*,T אנו בוחרים מסווג בסיסי *h*(*t*) מקבוצה של H מסווגים ושמים לו את המקדם α(*t*). בגרסה הפשוטה ביותר של האלגוריתם, H היא קבוצה סופית של סיווגים בינאריים מהצורה: *h* : **R***d* → {−1,1}, והלומד הבסיסי מבצע חיפוש ממצה על H בכל איטרציה.

הפלט של ADABOOST הוא פונקציה דסקרימיננטית הבנויה כהצבעות המשקלים המשוקללים של הסיווגים הבסיסיים.

הסימן של *f*(*T*)(x) בסוף משומש כמסווג הסופי של x.

האלגוריתם מתחזק פריסת משקלים w על נקודות המידע[[1]](#footnote-1).

המשקלים מאותחלים בצורה אחידה בשורת הקוד הראשונה, ומעודכנים בכל איטריצייה בשורות 7-10. המטרה של הלומד הבסיסי הוא למזער את השגיאה הממושקלת.



המקדם α(*t*)  של *h*(*t*)בשורה 5 נקבע כ: 





(2.4) המשקלים נשארים מנורמלים לכל אורך האלגוריתם כלומר, לכל t

כדי לראות זאת, ראשית נשים לב כי האיתחול מבטיח כי

לאחר מכן, בהנחה ש-

נקבל:

ב- (2.5) השמנו את המשקל המעודכן בשורה 8 ו 10, ב- (2.6) השתמשנו בעובדה כי  . ב- (2.7) השמנו את ההגדרה של ב- (2.2)

וההנחה האינדוקטיבית של

בשל כך שהמשקלים מנורמלים,  הוא מספר ממשי בין 0 ל 1. יתר על כך, אם  סגור תחת שלילה, אנו יכולים להבטיח כי

מאחר ואחרת ישתנה סימנו וישתמש ב במקום ב

ההשלכה הראשונה של  
היא ש תמיד אי-שלילי. המקדם עולה בצורה מונוטונית ככל ש פוחת  
  
שמשמעותו היא שככל שהמסווג הבסיסי  טוב יותר, כך "הצבעתו" לסיווג הסופי גבוה יותר.

ככל ששגיאת הבסיס,  שואף ל 0, כך המקדם  שואף לאינסוף.

בצורה פורמלית, מסווג בסיסי "חזק" עם  בצורה אוטומטית יקבל מקדם ששואף לאינסוף, כך שישלוט על הקומבינציה הליניארית

בעיקרון, מכך שמטרת ADABOOST היא למזער את שגיאת האימון,  משמעותו היא שהמטרה הושגה, כך שזה בסדר לעצור את האלגוריתם. בצורות מסויימות של ADABOOST זה יהיה שמיש להמשיך גם אם 

מצג 2.2: המקדם הבסיסי כפונקציה של שגיאת הבסיס . ב

ו ככל ש

ההשלכה השנייה של היא שכאשר המשקלים  מעודכנים בשורות 7-10, המשקל של נקודות לא מסווגות גדל (שורה 8), והמשקלים של נקודות שסווגו נכון פוחתות (שורה 10). ככל שהאלגוריתם מתקדם, המשקלים של הנקודות באופן תכוף לא סווגו נכון יטו להיות גדולות, שך שהמסווג הבסיסי יתמקד בנקודות ה"קשות" הללו.

לבסוף, על ידי הסתכלות על ההתפתחות של המחובר הראשון של (2.5), נוכל לשים לב כי השגיאה הבסיסית  של המסווג הבסיסי ה- t הוא 1/2, כך שלעולם לא נבחר את אותו מסווג בסיסי בשתי איטרציות רצופות.

  
זה אם  אזי . נתון זה מסופק ע"י רוב המסווגים הפרקטיים, ונניח זאת מעכשיו הלאה. אם  אינו סגור תחת השלילה ואיננו מצליחים למצוא מסווג בסיסי עם

נוכל להמשיך את הלולאה אך מאחר ו לכל, עדיף לנו לסיים בפועל.

**2.3 אנליזה בסיסית**

תחילה אנו מאחדים את שני המקרים בנוסחת עידכון המשקולות (שורות 7-10):



ב- (2.8) אנו משתמשים בהגדרה שב(2.3) של המקדם  ובעובדה שגם  ו  הינם שייכם לקבוצה {1,1-}. בשימוש של המשוואה המאוחדת (2.9), ההגדרה של (2.1) של הפונקצייה הדיסקרימיננטית הסופית  והאתחול האחיד של המשקלים בשורה 1, נוכל לבטא את המשקלים בצורה שאיננה רקורסיבית כ:



ראשית, הנוסחא הלא רקורסיבית הזו אומרת לנו שהמשקלים הינם פרופורציונלים ל

האקספוננט השלילי הינו השוליים הלא מנורמלים, שמוגדרים בצורה כללית עבור הפונקצייה הדיסקרימיננטית f ונקודה ((x,y



הכמות הזו תשחק תפקיד הכרחי באנליזה מבוססת שוליים של השגיאה המוכללת.

כעת, ברור כי:  
1) הסימן של השול  מסמן האם  מסווג נכון ע"י  או לא.

2) ככל ש גדול יותר, כך יותר "בטוח" המסווג של.  
אזי באופן אינטואיטיבי, היגיוני שהמשקל פוחת מונוטונית עם .

שנית, מ (2.7) אנו יודעים שהמשקלים מסתכמים ל1, כך שמהנוסחא הלא-רקורסיבית (2.10) זה מתקבל:



ע"י השוואה של (2.10) ו (2.12) אנו מקבלים באופן אוטומטי את משוואת המפתח



של תיאוריית ההתכנסות:

תיאוריה 1: בהנחה כי קיים קבוע חיובי  עבורו שגיאת הבסיס חסומה מלמעלה על ידו כך: 

(\*\*\* אולי יהיה טוב יותר. חסם תחתון על הקצה) עבור כל t שגיאת האימון

תהיינה 0 לאחר לכל היותר איטרציות.

הוכחה: ראשית, אנו חוסמים מלמעלה את שגיאת האימון ע"י:



ב(2.15) אנו משתמשים ב,(2.16) נראה ע"י (2.13), ב(2.17) אנו מיישמים את ההנחה כי  ו(2.18) מתקבל מ.

התיאוריה עוקבת לאחר מכן מהטבע הדיסקרטי של שגיאת האימון: לאחר שהחסם העליון יורד מתחת השגיאה חייבת להיות 0.

החשיבות העיקרית של תיאוריה 1 היא שהיא מביאה תשובה חיובית להשערת האצת הלמידה PAC: אפשרי לבנות לומד חזק (עם סיכון קרוב לאפס) ע"י שילוב מספר קטן של לומדים חלשים (עם סיכונים של מעט מתחת לחצי). מהירות ההתכנסות האקספוננציאלית של שגיאת האימון, מראה כי מספר הלומדים החלשים הם ~ ln n, משחק תפקיד חשוב בהוכחת ההשערה.

החסם העליון (2.16) מסביר את המטרה העיקרית של מזעור . בנוסף ברור שע"י מזעור , אנו לא ממזערים באופן ישיר את שגיאת האימון , רק ע"י החסם העליון.



נקרא הסיכון האקספוננציאלי הניסיוני. מזעור של חסמים עליונים קמורים במקום מזעור שגיאת האימון היא גישה נפוצה בלמידת מכונה משתי סיבות. הראשונה, מזעור

היא לעיתים קרובות פעולה קשה לחישוב, אפילו במחלקות פשוטות מבחינה פונקצינאלית. שנית, מזעור מונוטוני מפחית הפסדי שוליים אשר מוצדקים לעיתים קרובות ע"י ויכוחים הסתברותיים: אנו מקבלים שוליים גדולים יותר ושגיאות גנרליות טובות יותר ע"י "החלקת" השגיאה הניסיונית.

החסם העליון (2.16) ממחיש גם כן תכונה חשובה של ADABOOST: הטבע החמדי של בסיס האופטימיזצייה שלו. אכן, אנו לא ממזערים את הסכנה האקספודנציאלית באופן גלובלי מעל הקומבינציות הלינאריות של , אלא, בכל איטרצייה t אנו מוסיפים את הלומד הבסיסי  והמקדם שלו אשר הינו אופטימאלי בהינתן המסווג אשר נבנה עד כה.

פשרנות כללית זו של ADABOOST של ייעול הסיכון האקספוננציאלי באופן חמדני, הניב מספר הכללות והרחבות.

התנאי (2.14) בפועל חזק יותר ממה שהוא נראה: אפילו עבור קבוצת נתונים נתונה הוא דורש ששגיאת הבסיס תהיה תחומה מעל חצי עבור כל פריסת משקלים .

למעשה יכול להתרחש בקלות ש(2.14) עלול להיות לא מספק: כל מה שאנו צריכים זה קבוצת נתונים  שלא יכולים להיות מופרדים על ידי שום קומבינצייה ליניארית מעל 

במקרה זה, שגיאת האימון לא יכולה בשום פנים ואופן להגיע ל0, ושגיאת הבסיס  נוטה לחצי. בגבול, פריסת המשקלים היא כזו כך שלכל  שגיאת הבסיס הינה חצי.

מצד שני, אם ניתן להפרדה ע"י קומבינצייה ליניארית מעל , אזי אנו יכולים להוכיח כי (2.14) תמיד יכול להיות מסופק, ושהערך הממשי של  (וכך גם מהירות ההתכנסות) יכולים להיות קשורה להפרדת השוליים בין שתי המחלקות.

המונח  אשר מופיע ב(2.16) נקרא הקצה של מסווג הבסיס אשר מסומן ע"י .

בתחושה מסויימת, זוהי כמות טבעית יותר מאשר השגיאה  וזה יקל על הסימון במספר הצהרות ונגזרות. קל לראות ש:



כאשר (2.20) מתקבל מ  ו(2.21) נכון מאחר וגם  ו  שייכים לקבוצה {1,1-}.

הקצה של הבחירה הרנדומלית הוא 0 (בממוצע), אזי מונה כמה טוב הרבה יותר מאשר בחירה רנדומלית. מספר נוסחאות של האלגוריתם הבסיסי יכולים לפשט או יכולים להפוך להיות יותר "סימטריים" על שימוש בקצה במקום בשגיאה. לדוגמא, מקדם הבסיס (2.3) יכול להיות משוכתב כ:



  
  
המשקלים מעודכנים בשורות 7-10 מפושטים ל:

והסיכון האקספונדציאלי (2.13) יכול להיות מבוטא כ:



ההגדרה של השול (2.11) מספק אינטרפטצייה נוספת ל (2.21): הקצה הוא השול הממוצע הממושקל של המסווג הבסיסי . למעשה הקצה קשור לשול באופן דיסקרטי.

כהערה סופית, נשים לב כי החסם העליון (2.18) יכול להיות מחודד ע"י שימוש בקצוות  בדיעבד, במקום בגבול התחתון שלהם  מראש, אל:



אזי ADABOOST מבטיח להשיג 0 שגיאות אימון כאשר  עובר את .2ln(n)

1. כדי להימנע מבלבול, מעתה והלאה נקרא למסווגים הממושקלים

   *wi* מקדמים, ונשמור על המונח משקל עבור נקודות המידע α(*t*)  [↑](#footnote-ref-1)