# Ensemble Methods

*תקציר: במהלך ההכשרה, אסכם מספר נושאים "רוחביים" שאינם מתקשרים לפרק ספציפי. הנושא השלישי הוא ensemble methods . במסמך זה אסביר מה זה ensemble methods, למה משתמשים בזה ואילו שיטות קיימות.*

## רקע

מודל מבוסס ensemble יעשה שימוש בתוצאות של כמה מודלים (לרוב פשוטים) על מנת להגיע לתוצאה סופית טובה יותר מזו של אחד מהמודלים הנ"ל בנפרד. המודלים שמשמשים את מודל ה ensemble צריכים להיות שונים זה מזה בצורה כלשהי על מנת שלשילוב התוצאות יהיה איזשהו ערך. דוגמאות לכך יהיו שימוש באלגוריתמים שונים, חלקים שונים מהדאטה או היפר-פרמטרים שונים. היות וההחלטה מבוצעת על ידי מודלים פשוטים ושונים, אשר אומנו בצורות שונות, מודל ensemble לרוב יהיה בעל bias ו variance נמוכים. מצד שני, שימוש במספר מודלים מגדיל את הכוח החישובי ולעיתים את הזמן (במידה ואי אפשר למקבל את אימון המודלים) שלוקח לאמן מודל ensemble. בנוסף, יש חשיבות רבה לבחירת הדרך בה עושים ensemble (אילו מודלים, איך לאמן, איך לשלב את התוצאות וכו').

## שיטות מוכרות

ישנן כמה שיטות ensemble פופולריות, עליהן אפרט.

### Bagging

או Bootstrap Aggregating. בשיטה זו מבצעים bootstrapping לדאטה הקיים – שליפה אקראית של דוגמאות מהמידע המקורי (חזרה על אותה הדוגמה מותרת) על מנת לבנות כמה סטים של נתונים עליהם נאמן מודלים. כל מודל מאומן על סט אחר ובכך אנו יוצרים מודלים ששונים זה מזה. לבסוף מבצעים אגרגציה (למשל, הצבעה במקרה של סיווג או ממוצע במקרה של רגרסיה) של תוצאות המודלים. באיור 1 מוצגת אילוסטרציה של התהליך.

|  |
| --- |
| Image for post |

איור 1: תיאור של תהליך Bagging

יתרון יחסי של bagging על פני שיטות אחרות הוא שאפשר למקבל את בניית המודלים (הם לא תלויים אחד בשני).

Random forests הם סוג של bagging אבל במקום לדגום דוגמאות מהמידע אנו דוגמים עמודות (פיצ'רים) שלו.

### Boosting

בכל איטרציה נאמן מודל על המידע, כאשר ניתן חשיבות רבה יותר לדוגמאות שהמול מהאיטרציה הקודמת טעה לגביהן. באיור 2 מוצגת אילוסטרציה של התהליך.

|  |
| --- |
| Image for post |

איור 2: תיאור של תהליך Boosting

כמה מונחים מוכרים מהתחום:

* Adaptive Boosting: Adaboost, המודל הסופי מורכב מממוצע משוקלל של כלל המודלים שפותחו לאורך התהליך (בניגוד לתיאור לעיל, בו המודל הסופי הוא המודל האחרון בלבד).
* Gradient Boosting: כמו adaboost אבל השיטה למצוא המשקולות עושה שימוש ב gradient descent. Xgboost וחבריו שייכים לפה.

### Stacking

מודלים (במקרה זה מסוגים שונים) מאומנים על המידע שליש לנו (מודלי שלב 0) ומודל נוסף שמקבל כקלט את התוצאות של מודלי שלב 0 ומחזיר תוצאות סופית (מודל שלב 1). כלומר, משתמשים במודל כדי ללמוד כיצד לשלב את התוצאות של מודלים אחרים בצורה הטובה ביותר. אילוסטרציה שבצורה חד משמעית לא יוצרה על ידי בן אדם מתחום המחשבים (מי מתחיל לספור מ-1??) מוצגת באיור 3.

|  |
| --- |
| workflow |

איור 3: תיאור של תהליך Stacking

## ניסוי

הרצתי עץ החלטה, מודל bagging מבוסס על עצי החלטה ומודל stacking כאשר מודלי שלב 0 היו עץ החלטה, KNN ו SVC ומודל שלב 1 היה עץ החלטה (כל המודלים עם היפר-פרמטרים שנקבעו על ידי sklearn כברירת מחדל). התוצאות וזמני הריצה ב 10FCV מצורפים בטבלה 1 וגרף (בשביל הלוק) מצורף ב.

|  |  |  |
| --- | --- | --- |
| מודל | דיוק ממוצע | זמן ריצה כולל (שניות) |
| DT | 0.83 | 0.23 |
| Bagging | 0.90 | 2.03 |
| Stacking | 0.96 | 10.99 |

טבלה 1: סיכום ניסוי

|  |
| --- |
|  |

איור 4: גרף עמודות שמציג את הדיוק הממוצע של כל שיטה