למידה חישובית והמתודה המדעית

עבודה סמינריונית בקורס "אשנב לפילוסופיה של המדע"

מוגשת לפרופ' ימימה בן מנחם

מס' קורס: 87503

מגיש: רשף מאיר

ת.ז. 040097503

תאריך: 2/12/06

<u>תוכן העניינים</u>

מבוא 3

החקירה המדעית - רקע 4

למידה חישובית - רקע 6

מה מערכת לומדת לומדת? 6

מודלים פורמליים 6

אמת מוחלטת ואמת אמפירית 9

פופר 9

למידה חישובית 10

השערות נועזות ומורכבות של היפותזות 11

בעיית האינדוקציה 11

השערות נועזות כמענה לבעיית האינדוקציה 12

היפותיזות פשוטות כמענה לבעיית תת ההיקבעות 12

מושגים מקבילים בלמידה ובפילוסופיה 16

ביקורת 18

סיכום 21

נספחים – הגדרות, משפטים ודוגמאות 23

ביבליוגרפיה 27

מבוא

כאשר מדענים כגון אסטרונום, ביולוג, כימאי או אגרונום מבצעים מחקר על מנת לנסח חוק חדש או לגלות תופעה מסוימת, הם נדרשים לחשיבה יצירתית, איסוף של מידע ניסויי או אחר, ביצוע חישובים רבים, הסתמכות על הניסיון, למידה מטעויות ולעיתים ניצוץ של גאונות.

כאשר שרת הדואר האלקטרוני שלנו מחליט להשליך לאשפה מכתב מסוים, או כאשר מצלמת אבטחה מתריעה על תנועה חשודה, הם מריצים אלגוריתם שנכתב ע"י המתכנת - אך גם מתייחסים להחלטות הקודמות שביצעו ולומדים משגיאות. התחום העוסק בפתרון בעיות מסוג זה ידוע **כלמידה חישובית**.

האם באמת יש קשר בין תהליכים אלו, או שמא הדמיון הוא שטחי בלבד ונובע מאופן הניסוח? קיים לפחות מאפיין משותף אחד לשני תחומים נפרדים אלה, והוא השימוש באירועים מהעבר על מנת לדעת יותר על אירועים שיקרו בעתיד.

לאחר שאתן רקע בסיסי, אציג את בעיית האינדוקציה והקשיים הדומים שהיא מערימה על כל תהליך בעל אפיון דומה לתהליכים הנ"ל. לאחר מכן אטען כי חלק מן הרעיונות הפילוסופים שהועלו בניסיון להתמודד עם בעיה זו בתחום החקירה המדעית, קיבלו ניסוח פורמלי-מתמטי והם משמשים בצורתם החדשה את המהנדס המתכנן מערכת לומדת.

העבודה תתמקד בתורתו של הפילוסוף קרל פופר. להלן סכמת הקשר שאותו אציג בהרחבה בהמשך.

לפי פופר, על תיאוריה מדעית לעמוד בדרישות מחמירות על מנת שתתאפשר התקדמות לקראת האמת[1]:

1) על המדען לנסח את השערותיו על העולם **מראש**, ואז לבדוק אותן.

2) אלו צריכות להיות **השערות נועזות** שאין באפשרותן להסביר כל תצפית אפשרית.

כדי שמערכת ממוחשבת תוכל להצליח בסיווג וזיהוי של קלט עתידי:

1) על המהנדס **להבנות לתוך המערכת מראש** הנחות על העולם בו היא עתידה לפעול.

2) הנחות אלו צריכות להיות **חזקות** ע"מ שהמערכת תוכל ללמוד ביעילות.

מדוע אין אנו מבצעים תחילה תצפיות, ואז לומדים מהן? כפי שהסבירו יום ופופר, הבעיה היא חוסר היכולת להסיק באופן ודאי ממה שכבר ראינו למה שעתיד לקרות. יש אינסוף תיאוריות המתיישבות עם כל אוסף סופי של תצפיות, ויהיו תצפיות אלו גופים נופלים או מכתבי דואר אלקטרוני.

אם נניח הנחות מסוימות על העולם, נוכל לגזור מהן את התופעה שאנו מצפים שתקרה בתנאים מסוימים. אם התצפיות אינן מתיישבות עם הנחות אלה, נדחה אותן ונחליפן בהנחות אחרות.

בעבודתי אראה כיצד ניתן לתרגם מושגים כמו אמת, תצפיות והשערות למושגים פורמליים, וכיצד ניתן למדוד כמה השערה היא "נועזת" ולהשוות אותה עם אחרות. אראה כי גם מושגים נוספים כמו פשטות ואלגנטיות הנקשרים פעמים רבות לתיאוריות מדעיות, זכו לניסוח פורמלי והם רלוונטיים לפתרון בעיות בתחום הלמידה החישובית.

---

[1] אני מניח כאן שקיים קשר בין "תיאוריה אמיתית", "בעלת יכולת ניבוי" ו"בעלת יכולת הכללה". בהמשך העבודה ארחיב על הקשר בין מושגים אלה.

החקירה המדעית - רקע

בפרק זה אציג באופן כללי מאוד את החקירה המדעית והעקרונות המנחים אותה בעיני הוגים שונים. לאחר רקע היסטורי קצר אתאר את המתודה האמפיריציסטית, ואסיים בבעיית התיחום של פופר.

מקדמת דנא שרוי האדם בניסיונות להבין את הטבע מסביבו. עצם הישרדותו היה תלוי ביכולתו לענות, בין השאר, על השאלות הבאות:

האם ניתן לאכול מהצמח הזה בעל הפירות האדומים? ומה בנוגע לצמחים הדומים לו? האם המשחה המועילה לחתכים תיטיב גם עם כוויות? מתי ירד שוב גשם? מהו המיקום האופטימלי בנהר להשלכת החכה? ובאילו סימנים ניתן להיעזר על מנת לנווט חזרה הביתה?

כמובן שאין די בניסוי-וטעייה פשוטים ע"מ לענות על השאלות, שכן מגוון המצבים והווריאציות השונות לאותה בעיה הוא עצום. לאדם המסכן לא הייתה ברירה אלא להשליך ממצבים שכבר נתקל בהם אל מצבים חדשים, ופעמים רבות לטעות עקב כך.

האנשים שענו על שאלות אלה טוב יותר (או אולי בסמכותיות רבה יותר) מכל השאר זכו להוקרה רבה וסומנו כמכשפים וידעונים. תשובותיהם לא תמיד היו נכונות או מועילות, אך הם סיפקו לאנשיהם תחושת ביטחון מסוימת בעולם אקראי ומתעתע. האנשים ידעו כי יש הסבר לכל תופעה, גם אם הסבר זה היה לעיתים מיסטי ומעורפל, או כזה שאינו נותן כל אפשרות לחזות טוב יותר את העתיד.

אריסטו

במהלך הזמן החיפוש אחר ההסברים הפך למטרה, גם כאשר להסברים אלה לא היה כל שימוש פרקטי מיידי. הידיעה, או האמת, הפכה לעיקר. "כל בני האדם שואפים מטבעם אל הדעת. סימן לדבר התענוג הבא מן החושים. כי גם מבלי להתחשב בתועלת שבחושים, יש לנו תענוג מהם כשהם לעצמם..."[2]

בכך התעלה האדם מעל לשאר החיות, שגם הן התמודדו בדרך זו או אחרת עם השאלות שהצגנו קודם לכן. האדם בניגוד אליהן יכל להעביר הלאה את הידע והניסיון ובכך לאפשר את תחילתו של מה שקרוי החקירה המדעית[3].

אריסטו הפריד את החוכמה, או המדע, מן הניסיון לא באמצעות המבחן הפרקטי של ההצלחה, אלא באמצעות היכולת לספק הסבר אקספליציטי: "...הסימן של היודע והבלתי יודע הוא היכולת ללמד, ולפיכך סוברים אנו שהאומנות מדע היא באופן מדויק יותר מן הניסיון."[4]

במאות הבאות השאלות השתנו הוגים וחוקרים הוסיפו לעסוק בשאלת החקירה המדעית וכיצד עליה להיראות.

אמפיריציזם

בסביבות המאה ה 17, בעקבות דקארט ובייקון, התגבשה המתודה האמפיריציסטית השולטת במדע עד ימינו. זו ייחסה חשיבות עליונה לתצפיות לשם ניסוח תיאוריה מדעית.

עד למאה העשרים, רווחה בקרב האמפיריציסטים הגישה שהמדע מתקדם באמצעות איסוף תצפיות והסקה אינדוקטיבית של תיאוריה מתוכן[5]. תיאוריות אלה ניתן לאמת או להפריך באמצעות תצפיות נוספות. ההישענות על תצפיות אמפיריות היא שנחשבה לאבן הבוחן הקובעת האם תיאוריה תחשב למדעית.

---

[2] אריסטו, *המטאפיסיקה, ספר א'*, מיוונית: ח"י רות, תשנ"ח, עמ' 21.

[3] שם. עמ' 22-23.

[4] שם. עמ' 24.

[5] Markie, Peter, *Rationalism vs. Empiricism*, in Stanford Encyclopedia of Philosophy

היכולת להסביר מגוון רב של תצפיות נחשבה כמובן ליתרון של התיאוריה, שכן היא הקטינה את הסיכון להפרכתה, ומכאן הגדילה את הסיכוי לאמיתותה.

פופר

בתחילת המאה העשרים, הציע קרל פופר גישה שונה וחדשנית. עצם השימוש בתצפיות כקריטריון לעבודה מדעית לא סיפק אותו ולא איפשר להבחין,למשל, בין האסטרונומיה לאסטרולוגיה, שכן שתיהן מסתמכות על תצפיות אמפיריות.

בבואו לפתור את מה שקרא "בעיית התיחום" (problem of demarcation) בין תיאוריות מדעיות לכאלה שאינן, הציב פופר דרישות חדשות שעל תיאוריה לעמוד בהן כדי להיחשב מדעית[6]:

- עליה להיות ניתנת להפרכה.
- עליה "לאסור" כמה שיותר דברים, כלומר ככל שניתן לחשוב על תצפיות רבות יותר שיפריכו את התיאוריה, כך היא טובה יותר.
- באופן כללי תצפיות חיוביות אינן מאששות את התיאוריה. למבחן אמיתי יחשב רק ניסיון רציני להפריך אותה.
- במידה ועולה מהתיאוריה ניבוי שהופרך ע"י תצפית, אין "לתקן" אותה באמצעות הוספת הנחות נוספות.

המתודה המדעית עפ"י פופר היא הפוכה מהמתודה האמפיריציסטית המקובלת שהצגנו: המדע אינו מתקדם באמצעות איסוף תצפיות ושימוש בהן כדי להסיק תיאוריות, אלא באמצעות העלאת השערות (תיאוריות) וניסיונות להפרכה (ביצוע תצפיות ועימותן עם הניבויים מהתיאוריה).

כל תיאוריה, עפ"י פופר, עלינו לקבל באופן זמני בלבד, ועלינו להמשיך ולנסות להפריכה ולא לחפש דוגמאות שיאששו אותה. תיאוריה העומדת במבחנים קשים תוכל להמשיך ולשמש אותנו וערכה יעלה, אך לעולם אל לנו לקבל אותה כאמת מוחלטת.

אם אנסה לתמצת את דרישותיו למשפט יחיד, אזי תיאוריה מדעית טובה היא זו שמסבירה היטב את התצפיות הקיימות, וכמה שפחות דברים אחרים. ניסוח זה הוא הרלוונטי ביותר בהקשר שלפנינו, ועוד אחזור אליו בהמשך.

---

[6] הניסוח מבוסס על הרצאתו של קרל פופר "מדע: השערות והפרכות", מאנגלית: יורם נבון.

למידה חישובית - רקע

בחלק זה אציג את סוג הבעיות אותן מנסים לפתור ע"י מערכות לומדות. לאחר שאגדיר מספר מושגים בסיסיים בתחום, אסקור כמה סוגים של בעיות, ובייחוד בעיות[7] קלסיפיקציה. אנסה להראות מדוע מסגרת פורמלית זו מתאימה למגוון רחב של בעיות שימושיות. חלק מבעיות אלה הן טכנולוגיות גרידא, אך בקלות ניתן לשאול במסגרת זו (אם כי לא כל כך קל לענות!) שאלות רחבות יותר, דוגמת הבעיות המדעיות והיומיומיות שהצגנו בחלק הקודם.

על סמך הגדרות אלה אציג בחלקים הבאים את תהליך הלמידה עצמו ואת התנאים בהם הוא קורה.

מה מערכת לומדת לומדת?

למידה חישובית היא שם כולל לאוסף רחב של בעיות, אפליקציות, מודלים ואלגוריתמים שהמשותף לכולם הוא הניסיון להסיק מידע חדש מתוך מידע מסוים הנמצא בידינו.

אלו הן שתי סוגיות מרכזיות בהן עוסקת למידה חישובית:

א. שימוש במידע קיים על מאפיינים של עצמים ותופעות בעולם כדי לנחש או לנבא מאפיינים של עצמים אחרים, או תחת תנאים אחרים. בעיות אלה נקראות בעיות **טרנסדוקציה**. אין הן מחייבות **הבנה** כלשהי של מצבים או תהליכים בעולם. די לנו שהניבוי הוא מוצלח, ואין אנו מבקשים מן המערכת הלומדת לתת הסבר שנוכל להבין, או הסבר כלשהו.

ב. שימוש במידע הקיים כדי לקבל **כללים**. בעיות אלה מוכרות כבעיות **אינדוקציה**. כאן אנו מבקשים מהמערכת הלומדת לתת לנו הסבר אקספליציטי שבעזרתו נבין את המידע שברשותנו, ונוכל ליישמו במקרים נוספים.

כפי שניתן לראות, פתרון בעיה כלשהי בדרך ב' גורר גם שיש פתרון מסוג א': אם יש ברשותנו הסבר כללי לתופעה, ניתן להפעיל אותו על עצמים ותנאים ספציפיים כדי לקבל ניבוי. בדיוק כפי שהסקה של פונקציה כלשהי מתוך ערכיה הנתונים במספר נקודות, מאפשרת לנו לשערך אותה גם בנקודות אחרות.

בסיטואציות פרקטיות, מה שנדרש הוא בד"כ לתת מענה למקרים ספציפיים, כלומר לפתור בעיות מהסוג הראשון. למרות זאת, מרבית המחקר בתחום עוסק בבעיות מהסוג השני, וגם עבודה זו תתמקד בכך[8].

מודלים פורמליים

להלן כמה הגדרות ומודלים פורמליים המשמשים אותנו בלמידה חישובית, ואשתמש בהם גם בעבודה זו.

ההגדרות וחלק מן הדוגמאות מבוססות על הכתוב בספר "An Introduction to computational learning theory" / Michael J. Kearns & Umesh V. Vazirani

---

[7] המושג "בעיה" משמש אותנו גם לציון בעיה מדעית/הנדסית ספציפית אותה ברצוננו לפתור (למשל זיהוי פרצופים, אפיון הקשר בין שני גנים וכו'), אך גם בהקשר הרחב יותר ("בעיית האינדוקציה", "יש לנו בעיה לפתור את המשוואות הבאות..."). אנסה להשתמש במילים חלופיות עבור השימוש השני, או להבהיר את המשמעות באמצעות ההקשר.

[8] למעשה המונח "בעיות טרנסדוקציה" הוא חדש יחסית, והוצג רק בשנות ה90 ע"י Vapnic. הוא עצמו טען שהבעיות מהסוג הראשון מוצגות באופן שגוי ("the ill-posed problem")
Vapnic, Vladimir, *Statistical Learning Theory*, Wiley-Interscience (1998), pp. 339-343

ההגדרות והמשפטים מוצגים בצורה מפושטת, כאשר את המרכזיים ביניהם ניתן למצוא בצורתם הפורמלית בנספחים לעבודה זו.

מודל פורמלי

מודל פורמלי הוא תיאור של:

1) אופן הצגת נתוני הבעיה
2) אופן הצגת הפתרון
3) אופן ההערכה של פתרון או השוואה בין פתרונות שונים.

בנוסף המודל עשוי להכיל הנחות יסוד על העולם, ומגבלות על אופן הפתרון.

המודל אינו מתאר בעיה ספציפית – ניתן לתאר בעיות רבות באותו מודל, ולעיתים להציג בעיה מסוימת בכמה מודלים שונים. המודל אינו מתאר את הדרך לפתרון, ואינו מבטיח שניתן למצוא פתרון כלשהו.

המודל שאתאר הוא מודל ה PAC לקלסיפיקציה בינארית, בו גם אשתמש בהמשך.

קיימים מודלים נוספים, אך הם מחוץ לתחום של עבודה זו. עם זאת אציין כי את המסקנות שאציג בסיכום ניתן להחיל גם על חלק מהמודלים האחרים.

הגדרות כלליות

דוגמא (instance)**:** הייצוג הפורמלי של פריט ע"י וקטור דיגיטלי. בלמידה חישובית מייצגים כל דבר הלקוח מ"העולם האמיתי" ע"י וקטור, כלומר ע"י סדרה של מספרים. זאת בין אם מדובר בתוצאות סקר,תמונה, מידע גנטי, משפט בצרפתית או כל דבר אחר.

אופן ה"תרגום" של אוביקט ממשי (או אבסטרקטי) לווקטור דיגיטלי *אינו* נחשב לחלק מן הבעיה, ולא אעסוק בו בחלקים הבאים. פשוט מניחים שנעשה שימוש בתרגום כזה, שהווקטור אכן מייצג את הפריט האמיתי היטב, וכן שמדד המרחק[9] שמשתמשים בו משקף נכונה את מידת הדמיון בין הפריטים במציאות.

נסמן כל דוגמא ב x, את אוסף כל הדוגמאות האפשריות ב X, את הקבוצה (הסופית) של דוגמאות נתונות ב S, ואת גודלה ב m.

פילוג: התפלגות על פני X המתארת את הסיכוי "להיתקל" בכל אחת מן הדוגמאות בעולם. מסומן D.

היפותזה : נקראת גם "השערה" ו"מושג" (concept). פונקציה מכל קלט אפשרי לפלט אפשרי. מסומנת בד"כ כפונקציה מעולם הקלט X לעולם התשובות Y.

כאשר לכל דוגמא יש תשובה דטרמיניסטית[10], נהוג לסמן את התשובה c(x)=y, ולקרוא לה "פונקצית המטרה".

כאשר Y רציפה (גובה, זווית, מיקום במרחב וכו') זוהי בעיית **רגרסיה**.

כאשר Y מכילה מספר סופי של ערכים (דוגמאות: סוגי דם, אותיות באנגלית וכו') זוהי בעיית **קלסיפיקציה**.

בפרט אם Y מכילה 2 ערכים בלבד (כן/לא, 0/1, תקין/לא תקין וכו') זוהי בעיית קלסיפיקציה בינארית.

לשם הפשטות, נעסוק מעתה והלאה בבעיות קלסיפיקציה בינארית, ובמקרים בהם קיימת פונקצית מטרה.

---

[9] אם חושבים על כל ווקטור מייצג כעל נקודה במרחב בעל n ממדים, דוגמא למדד דמיון טבעי בין דוגמאות הוא המרחק האוקלידי בין הנקודות המייצגות אותן. זהו אכן מדד נפוץ, אם כי ייתכנו גם מדדים אחרים.

[10] במקרה הכללי h אינה בהכרח פונקציה דטרמיניסטית. הפונקציה h עשויה לתאר גם את ההסתברות לקבלת כל y אפשרי בהינתן x. דוגמאות לבעיות דטרמיניסטיות והסתברותיות מופיעות בנספח ב'.

היפותזה במסגרת זו של בעיות קרויה גם **מסווג** (classifier), או **מפריד** (separator). זאת מכיוון שהיא מפרידה את מרחב הדוגמאות X לשני חלקים – דוגמאות עבורן Y=1, ודוגמאות עבורן Y=0.

מחלקת מושגים (concept class)**:** קבוצה כלשהי (סופית או אינסופית) של היפותזות. עשויה להכיל או שלא להכיל את פונקצית המטרה. מסומנת כ- C.

מחלקת מושגים בינארית : מחלקה בה כל ההיפותזות הן מפרידים. כלומר על כל דוגמא ההיפותזה משיבה "כן" או "לא".

שגיאה של היפותזה: ההסתברות שההיפותזה h תטעה על דוגמא x שנדגמה באקראי מהעולם X. מסומנת err(h)[11]. כדאי לציין שבמקרים אמיתיים רבים, ייתכנו אינסוף דוגמאות שונות בעולם (למשל מספרים רציפים), ולכן לא ניתן לחשב את err(h) בדיוק.

מודל ה PAC [12]:

הקלט – אוסף S של m דוגמאות ותשובותיהן הנכונות ("מדגם").

הפלט – היפותזה $\hat{h}$.

אופן הערכה - ככל ש $err(\hat{h})$ קטנה יותר, כך $\hat{h}$ טובה יותר.

מגבלה – ההיפותזה $\hat{h}$ לקוחה מתוך מחלקת היפותזות C *שנקבעה מראש*.

הנחות יסוד – כל הדוגמאות ב S נדגמו i.i.d [13] לפי D.

בעיה ניתנת ללמידה במודל ה PAC, אם ניתן לקבל היפותזה טובה על סמך מדגם בגודל סביר.

האם כל בעיה במודל הPAC ניתנת לפתרון? ננסח במודל בעיה שפתרונה יסייע למצבנו הכלכלי:

x הוא ייצוג של רשומות המסחר בבורסה הניו-יורקית במשך יום בודד.

S יהיה אוסף של רשומות המסחר ב 1000 הימים האחרונים. לכל רשומה מצורף הנתון "האם המניה של IBM עלתה ביום שלמחרת?"

h היא פונקציה מרשומה x לתשובה "כן" או "לא".

C היא אוסף **כל** הפונקציות h הנ"ל.

אם בעיה זו ניתנת ללמידה, הרי שביכולתנו לקבל פונקציה המנבאת את התנהגות מניית IBM.

נחזור לשאלה זו מאוחר יותר (ואולי נוכל לפרוש מהעבודה לצמיתות...)

---

[11] פורמלית: $err(h) = \Pr_{x\sim D}[h(x) \neq c(x)]$. בספרות מופיעה גם כ Risk(h).

[12] הגדרה פורמלית מופיעה בנספח א'.

[13] i.i.d – independent & indentically distributed (למשל הטלות מטבע בלתי תלויות)

אמת מוחלטת ואמת אמפירית

לפני שאעבור להצגת הבעיות השונות בחקירה מדעית ובלמידה חישובית, ראוי להקדיש כמה שורות לבחינת המטרה המוצהרת של כל אחת מהן.

נאמר שיש בידינו היפותזה/תיאוריה כלשהי, בכל תחום שהוא. שאלה מתבקשת היא "האם ההיפותזה נכונה?", או לחלופין "עד כמה ההיפותזה נכונה/שגויה?". גם אם אתעלם לחלוטין מגישות רלטיביסטיות ומתורות של משמעות (מה שאכן אעשה), עדיין ניתן לענות על השאלה בשתי דרכים שונות לפחות.

לפי הגישה האינסטרומנטליסטית שפופר מייצג[14], המדד היחיד לנכונותה של תיאוריה הוא הצלחתה לעמוד במבחנים אמפיריים. אם מובטח לנו בדרך כלשהי שניבוייה מתממשים תמיד, אזי היא נכונה. בדיקה כזו אמנם אינה בידינו, אך ניתן לכמת או להשוות בין תיאוריות באופן זה: אם ניבוייה של תיאוריה אחת מוצלחים יותר באופן גורף מאלה של תיאוריה אחרת, אזי היא טובה יותר.

ברם, עבור מדענים ופילוסופים רבים, אין די בסטטיסטיקה יבשה: היפותזות, גם באותו תחום, עשויות להתייחס לסיטואציות שאינן חופפות בדיוק ולכן קשה להשוות ביניהן. מעבר לכך יש הדורשים מן התיאוריה תכונות נוספות – למשל שתהיה פשוטה[15], או אסתטית, או שתספק לנו הסבר אקספליציטי שיאפשר לנו גם להבין את התופעות, ולא רק לנבא[16]. לפי גישות אלה, יתכן שנעדיף תיאוריה אחת על אחרת גם אם הראשונה אינה עולה על השניה מבחינת יכולת ניבוי, או אף נופלת ממנה.

בהמשך הפרק נראה כיצד באה לידי ביטוי הגישה האינסטרומנטליסטית אצל פופר וביישומים של למידה חישובית, ואילו מאוחר יותר נבדוק האם יש קשר בין למידה חישובית גם לגישות האחרות.

פופר

פופר מגדיר מושג של "התאמה" (fitness) ומאוחר יותר "אמיתיות" (verisimilitude) של פסוק או תיאוריה[17].

על מנת לעשות זאת, הוא מגדיר תחילה את "תוכן האמת" Ct ו"תוכן השקר" Cf של התיאוריה כקבוצת כל הפסוקים האמיתיים והפסוקים השקריים הנובעים ממנה, בהתאמה.

לאחר מכן הוא מגדיר את ה"אמיתיות" Vs של תיאוריה כך:

$$Vs(T) = Ct(T) - Cf(T)$$

כאשר אנו עוסקים בשתי תיאוריות הניתנות להשוואה, נעדיף את T1 על T2 אם $Vs(T1) > Vs(T2)$. תיאוריה טובה יותר היא זו שנותנת יותר ניבויים נכונים, ופחות ניבויים שקריים.

גם אם ניתן לפרט את כל הניבויים מתיאוריה מסוימת בדרך קומפקטית כלשהי, עדיין לא ניתן לבדוק את כולם. לכן עלינו לבצע מבחנים אמפיריים ולבדוק את התיאוריה על דוגמאות, אבל על אילו דוגמאות? ברור שאם נבדוק שוב ושוב על אותה הדוגמא, התוצאה לא תחזק או תחליש את התיאוריה.

לפי פופר עלינו לבדוק מקרים "קשים", כלומר כאלה שבהם היינו מצפים שהתיאוריה תיכשל.

גישה אחרת היא שעלינו פשוט להשתמש בכל המקרים שלפנינו, מבלי לבחור או לסנן אותם[18].

---

[14] אף שנטען כי גישה כזו סותרת את קיומה של אמת אובייקטיבית, פופר ניסה ליישב בין התפיסות. http://en.wikipedia.org/wiki/Instrumentalism

[15] גישה המוכרת בתור "התער של אוקהם". Baker,Alan, *Simplicity*, in Stanford Encyclopedia of Philosophy

[16] זוהי הגישה של אריסטו שהצגתי ברקע לעבודה.

[17] Popper, Karl, *Conjectures and Refutations,* London, 1973, pp. 231-235

למידה חישובית

היפותזות h1 ו h2 תמיד ניתנות להשוואה, כיוון ששתיהן פונקציות מאותה הצורה. נרצה לבחור בהיפותזה הקרובה יותר לאמת, אך מהי ה"אמת"?

אם רצוננו בתמונה המלאה אזי ה"אמת" היא הפילוג המשותף D על X^xY. נוח יותר להסתכל על פילוג זה כעל שתי רשימות[19]:

- הראשונה מתארת את ההסתברות לקבלת כל דוגמא אפשרית x in X.
- השנייה מתארת את ההסתברות לקבלת התשובה "1", בהינתן הדוגמא.

במקרים פשוטים בהם אפשרי סיווג מושלם[20] (קיימת פונקצית מטרה), הערכים ברשימה השנייה הם תמיד 0 או 1. במקרה כזה נאמר שכל דוגמא היא "שלילית" או "חיובית" בהתאמה, והאמת מיוצגת על ידי פונקצית המטרה c(x). לפי הגדרה זו err(h), שהוגדרה להיות ההסתברות של h לטעות על דוגמאות בעולם, מייצגת עבורנו את מרחקה של h מהאמת. לכן נעדיף את h1 על h2 אם"ם $err(h2) > err(h1)$.

נשים לב שחלק מהגדרת האמת היא ההסתברות להופעת כל דוגמא. גם אם הסתברות זו לא ידועה מניחים כי היא קיימת[21].

נותר לנו לקשר מושג זה של "אמת" ליכולת הכללה: בהינתן דוגמא חדשה x_{m+1}, הסיכוי של h לתת עבורה את התשובה הנכונה שווה בדיוק ל err(h)[22].

הערכה מעשית של היפותזה

נאמר שבידינו היפותזה h כלשהי. ייתכן שהשגנו אותה באמצעות אלגוריתם, ניחוש, או שקיבלנו אותה ממקור אחר. איך נעריך אותה? לכאורה זה פשוט, הלוא הגדרנו את err(h) באופן פורמלי. אך הגדרנו את השגיאה באמצעות **הסתברות**, כלומר באמצעות הפילוג D. בבעיות הקשורות לתופעות בעולם האמיתי, הפילוג "האמיתי" בד"כ אינו ידוע.

בהמשך אגדיר את השגיאה האמפירית של היפותזה ונראה תחת אילו תנאים היא מספקת לנו קירוב של השגיאה האמיתית.

---

[18] גודמן מציב דרישה זו במאמרו The new riddle of induction (1983) (פרק 3), אם כי בהקשר שונה.
[19] במקרה ש X או Y רציפות לא ניתן להשתמש ברשימות, אך לצורך ההמחשה נסתפק בכך.
[20] דוגמאות לסיווג מושלם ולא מושלם בנספח ב'.
[21] דוגמא להשפעת הפילוג על הערכת ההיפותזות בנספח ג'.
[22] זאת תחת ההנחה שהדוגמא החדשה נדגמה אף היא i.i.d מתוך הפילוג D.

השערות נועזות ומורכבות של היפותזות

בפרק זה אפתח את הרעיון המרכזי עליו בנויה העבודה. תחילה אציג בקצרה את בעיית האינדוקציה והשלכותיה על החקירה המדעית ועל למידה חישובית. לאחר מכן אראה כיצד מתמודדים עם השלכות אלה בתכנון מערכות לומדות על פי העקרונות שטבע פופר.

בעיית האינדוקציה

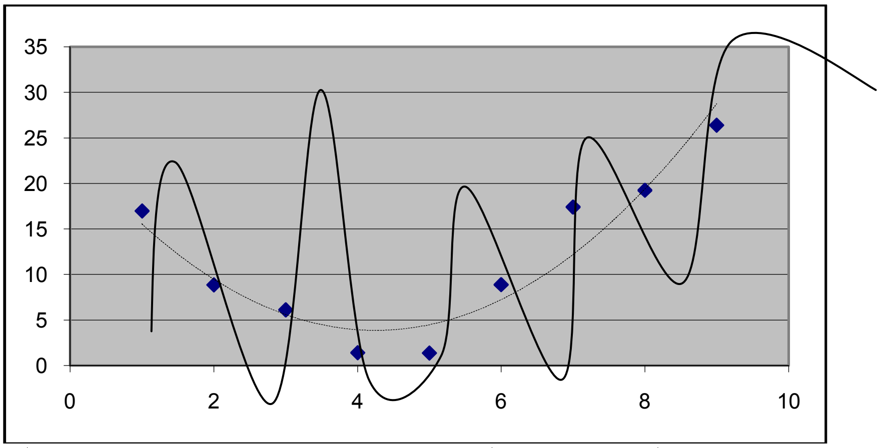
הבעיה היסודית במדע היא בעיית האינדוקציה. כפי שנוסחה על ידי דיוויד יום - מה שהיה בעבר אינו מטיל שום מגבלות על העתיד[23]. מופע בולט שלה הוא בעיית תת-ההיקבעות המעשית: בהינתן אוסף סופי של תצפיות ישנם אינסוף הסברים המתיישבים עם כולן. איך נמצא את ההסבר בעל יכולת החיזוי/ההכללה הטובה ביותר? (או יכולת הכללה כלשהי?)

אספקט נוסף של הבעיה הוא רגרסיה אינסופית – אנו משתמשים באינדוקציה כדי להסיק מה יקרה במצבים דומים למצבים אחרים, אך מושג הדמיון שאנו משתמשים בו גם הוא תלוי בניסיון ונלמד באופן אינדוקטיבי. בעיה אחרונה זו אינה קיימת במסגרת של למידה חישובית אותה הצגתי, וזאת מכיוון שהיא מתעוררת בשלב של תרגום הקלט מהעולם האמיתי לייצוג הממוחשב שלו. כזכור, שלב זה כולל גם את בחירת מדד הדמיון בין הדוגמאות, ואינו חלק מבעיית הלמידה החישובית כפי שהגדרתי אותה.

לעומת זאת, בעיית תת-ההיקבעות מהווה אתגר רציני בפני המהנדס הכותב מערכת לומדת. קל מאוד לכתוב אלגוריתם המוצא היפותזה המתיישבת עם כל N הדוגמאות שהתקבלו עד כה. למשל בבעיית רגרסיה (אין קשר לרגרסיה אינסופית) תמיד יש פולינום ממעלה N העובר בכל הנקודות[24]. באופן דומה בבעיית קלסיפיקציה תמיד יש פונקציה המסווגת נכון כל מדגם סופי בגודל N ע"י טבלה בגודל N.

למעשה יש אינסוף פונקציות טריוויאליות כאלה[25], אך אלו בדרך כלל חסרות כל יכולת הכללה.

לעומת זאת קיימות פונקציות "פשוטות יותר", שנראה כי הן מתארות את המידע באופן מקורב (איור 1), וגם שיאפשרו לנו להכליל לדוגמאות שטרם ראינו. איך בוחרים פונקציה "מתאימה"?



איור 1. הנקודות השמנות הן הייצוג של דוגמאות שהתקבלו ע"י תצפית. הקו הרציף "מסביר" אותן באופן מושלם.
הקו המקווקו הוא פונקציה הנראית פשוטה יותר (פרבולה) העוברת דרך הנקודות באופן מקורב בלבד. באיזו פונקציה

---

[23] יום, דיוויד, *מסכת טבע האדם*, ספר 1, הוצאת מאגנס, תשי"ד.
[24] http://mathworld.wolfram.com/LagrangeInterpolatingPolynomial.html
[25] כל פונקציה המסווגת נכון את n הדוגמאות שהופיעו, ונותנת ערכים שרירותיים לכל הדוגמאות שלא הופיעו.

נשתמש אם נרצה לנחש את תוצאת התצפיות הבאות?

<u>השערות נועזות כמענה לבעיית האינדוקציה</u>

למרות שיום טען שהאינדוקציה אינה תקפה, במישור הפרקטי מדענים הגיעו להישגים מרשימים בשיטות אינדוקטיביות. תיאוריות שהושגו באמצעים אלה הצליחו פעמים רבות לא רק להסביר תהליכים בדיעבד, אלא גם לנבא את התוצאות של התרחשויות עתידיות. חוקי קפלר שהושגו באמצעות מדידות שימשו את האסטרונומים בהצלחה תקופה ארוכה. כך גם נוסחאות וקבועים רבים בפיסיקה, כימיה ותחומים נוספים.

תיאוריות אחרות שהסתמכו על אינדוקציה ותצפיות זכו גם הן להצלחה, גם אם ניבוייהן היו פחות מוצלחים, מכיוון שהצליחו לספק הסברים משכנעים. כלפי תיאוריות אלה, בעיקר בתחומים של מדעי החברה, נטען פעמים רבות שהן אינן מדעיות מספיק.

פופר גרס שפתרון בעיית האינדוקציה טמון בפתרונה של **בעיית התיחום.** הוא סיפק קריטריונים ברורים מתי תיאוריה יכולה להיחשב כתיאוריה מדעית, אותם הצגתי בפרק 1.

כיצד הצבת קריטריונים אלה פותרת את בעיית האינדוקציה?

התהליך הלוגי העיקרי אצל פופר הוא הדדוקציה ולא האינדוקציה. אראה כיצד תהליך זה פועל על תיאוריה מסתכנת (שניתן להפריכה) ועל תיאוריה שאינה מסתכנת[26]:

- מלכתחילה לשתי התיאוריות סבירות זהה ונמוכה, שכן הן התקבלו בתהליך לא תקף כלשהו (למשל באינדוקציה).
- כאשר מעמתים תיאוריה מסתכנת עם תצפיות, היא עשויה להצליח להסבירן או להיכשל בכך. אם היא מצליחה אזי עולה הסבירות לאמיתותה, שכן הייתה יכולה גם להיכשל. לפיכך, אם נשתמש בתיאוריה כעת כדי לנבא תוצאות של תצפיות נוספות, יש סבירות גבוהה לניבוי טוב.
- כאשר מעמתים תיאוריה שאינה מסתכנת עם תצפיות, היא לרוב תצליח להסבירן, אך הצלחה זו אינה מעלה כלל את הסבירות לאמיתות התיאוריה. זאת ממש כפי שאין טעם לבחון תלמיד על שאלה שברור מראש שהוא יודע – תשובתו לא תשכנע אותנו שהוא חכם יותר ממה שחשבנו קודם. לכן, אם עתה נשתמש בתיאוריה לצורך ניבוי, אין שום יסוד להניח כי הניבוי יהיה מוצלח.

מכיוון שאנו מתייחסים לאמיתות של תיאוריה כאל היכולת שלה לספק ניבויים מוצלחים בסבירות גבוהה, נעדיף תיאוריות מסתכנות.

<u>היפותזות פשוטות כמענה לבעיית תת ההיקבעות</u>

אנסה לבנות את הטענה הפורמלית על התנאים ללמידה סטטיסטית בצורה דומה לצורה בה הצגתי את הטיעון של פופר.

לשם כך אזדקק להגדרה נוספת.

[26] צירפתי ניסוח פורמלי יותר של טענה זו באמצעות הסתברות מותנה. ראה נספח ד'.

שגיאת הכללה ושגיאה אמפירית

בפרק הקודם הגדרנו את השגיאה של היפותזה[27] להיות הסיכוי שלה לטעות על דוגמא מקרית. ניתן לראות ערך זה כערך המשלים ל"אמיתיות" של פופר.

err(h) היא ההסתברות לטעות על הדוגמאות הבאות, ולכן נקראת גם "**שגיאת ההכללה**".

כדי להראות שלהיפותזה h יש ערך "אמיתיות" גבוה, נרצה לחסום את err(h) מלמעלה.

סימנו ב $S = (x_1,x_2,..,x_m)$ אוסף כלשהו של דוגמאות.

נגדיר את **השגיאה האמפירית** (על מדגם ספציפי) להיות החלק היחסי מהמדגם S עליו h טועה[28]. מושג זה מקביל ליכולת של תיאוריה להסביר תצפיות קיימות.

נקודה עדינה וחשובה: כל היפותזה h זהה לחברתה מבחינת היכולת ההסברית[29], אין היפותזות "חזקות" ו"חלשות", ומשום כך לא ניתן להקביל בין היפותזה בודדת לתיאוריה במובן המדעי. אולם אם ניקח **מחלקת היפותזות** אזי ההקבלה אפשרית:

אם נסתכל על מחלקות בגודל סופי, אזי הגיוני שמחלקות גדולות יוכלו להסביר יותר מדגמים, ולכן יהיו קשות יותר להפרכה. עבור מחלקות קטנות ניתן בקלות למצוא מדגם שלא יוכלו להסביר ולהפריך אותן, בדומה ל"השערות הנועזות" של פופר. בחירת מחלקה קטנה כמוה כהשערת השערה חזקה.

כיצד אם כן מודדים "חוזק" של מחלקה בגודל אינסופי[30]? יש לכך מדדים שונים, קומבינטוריים, אנליטיים, וסטטיסטיים. כמו גודל של מחלקות סופיות, ככל שמדדים אלו גבוהים יותר, כך ההשערה חלשה יותר. בהמשך אציין כמה מדדים כאלה ואפרט את חשיבותם.

תהליך טיפוסי של למידה חישובית סטטיסטית

ניתן לנסח את תהליך הלמידה הסטטיסטית באופן דומה לניסוח תהליך החקירה המדעית עפ"י פופר:

כפי שראינו במודל ה PAC, בשלב הראשון המערכת מקבלת מחלקת היפותזות מוגדרת C. זוהי המקבילה לתיאוריה הנמצאת תחת מבחן.

כעת המערכת מקבלת אוסף של דוגמאות (=תצפיות) S ומנסה "להסביר" אותן. ההסבר שהיא מספקת הוא היפותזה בודדת h מתוך המחלקה שהמערכת מחזירה לנו.

מדוע זהו הסבר? כזכור h היא פונקציה מדוגמאות לתשובות. אם היא תמיד נותנת, ותמשיך לתת, את התשובה הנכונה אז די לנו בכך.

עתה הגענו לנקודה המרכזית – נאמר ש h אכן מסבירה היטב את המדגם S, האם תוכל לנבא גם את התשובות של דוגמאות נוספות?. שאלה זו שקולה לשאלה האם הצלחנו לקבל היפותזה עם err(h) קטן.

---

[27] עד סוף הפרק אבדיל בין "היפותיזה", שהיא פונקצית סיווג בינארית ספציפית, ל"תיאוריה" במובן הרחב יותר.

[28] פורמלית: $\check{e}r(h,S) = \frac{1}{m} | \{ i : h(x_i) \neq c(x_i) \} |$

[29] אם נניח שהתשובות הן אקראיות, כלומר שמראש יש הסתברות זהה לכל תוצאה של המדגם, אזי תוחלת השגיאה האמפירית של כל היפותזה היא תמיד ½.

[30] הסבר על מחלקות אינסופיות של היפותזות בנספח ה'.

הקשר בין מורכבות ויכולת הכללה

הבסיס התיאורטי לתהליך הלמידה הסטטיסטית שהצגתי הוא האפשרות להוכיח את קיומם של חסמים מהצורה הבאה:

עבור מדגם S גדול מספיק, אי השוויון:

$$| err(h) - e\check{r}r(h,S) | < COMPLEXITY (C)$$

מתקיים בהסתברות גבוהה

אסביר כיצד באה לידי ביטוי מורכבות מחלקת ההיפותזות שבחרנו (C), שאליה משתייכת h:

במחלקות עם מורכבות נמוכה המרחק בין השגיאה האמפירית לשגיאת ההכללה הוא קטן, ולכן נצפה שתיאוריה (מחלקת היפותזות) שמסבירה היטב את המדגם תספק לנו היפותזה עם שגיאת הכללה טובה. ככל שנבחר מחלקה פשוטה יותר, כך נוכל להעריך את השגיאה האמיתית טוב יותר.

בהסתמך על אי השוויון הנ"ל אבנה טיעון דומה לזה של פופר שהצגתי בחלק הקודם:

- נתונות שתי מחלקות של היפותזות: C1 עם מורכבות גבוהה, ו C2 עם מורכבות נמוכה.
- C2 עשויה להחזיר היפותזה h שמסבירה היטב את S (עם שגיאה אמפירית נמוכה), או לא להצליח להסביר את S. אם היא הצליחה, אזי מובטח לנו (בסבירות גבוהה) ש h תספק ניבויים טובים גם בהמשך. אם היא נכשלה אזי מובטח לנו שאף היפותזה ב C2 לא יכולה להסביר את אוסף התצפיות S, ולכן גם לא לספק ניבויים טובים[31]. כלומר התיאוריה ש C2 מייצגת הופרכה.
- בשל מורכבותה של המחלקה C1, היא כמעט תמיד תצליח להסביר את S, אך לא נוכל לומר דבר על יכולת הניבוי של ההיפותזה שהיא מחזירה, כי יתכן שהשגיאה האמפירית של h קטנה בהרבה משגיאת ההכללה שלה.

בנוסחה זו טמונה אף השיטה למצוא היפותזה טובה:

- לבחור מחלקה "פשוטה" (להקטין את המורכבות)
- בתוך מחלקה זו לחפש את ההיפותזה הטובה ביותר (להקטין את השגיאה האמפירית)
- אם השגיאה האמפירית גבוהה, נדחה את ההשערה (נחליף את המחלקה C)

ובכן, כעת ניתן לחזור לבעיית ניבוי ההתנהגות של מניית IBM. תחילה עלינו להניח כי אכן קיים "חוק" כלשהו המאפשר חיזוי מדויק מספיק של ערך מנייה על סמך רשומות המסחר. גם אז, אותו מהנדס החולם להתעשר צריך תחילה למצוא מחלקת היפותזות שתהיה פשוטה מספיק כדי שיוכל לבחון אותה. כנראה שאין זו משימה פשוטה למצוא מחלקה כזו שגם מכילה היפותזות טובות, שאם לא כן, מדעני המחשב היו נוטשים את האוניברסיטאות בהמוניהם לטובת השוק המבטיח. ובכל זאת, יש כאלה המנסים את מזלם[32].

---

[31] זה נובע מכך ש h ממזערת את השגיאה האמפירית. ראה נספח ז'.

[32] למשל ניסיון ללמוד את התנהגות שוק המניות בקראצ'י: Burney et al. (2004)

איך מחשבים מורכבות של מחלקת היפותזות?

Vladimir N. Vapnic היה הראשון להציג מושג פורמלי של מורכבות מחלקת היפותזות כבר בשנות ה60 המאוחרות. מושג זה, המוכר בשם "ממד VC", איפשר בעיקר הבחנה בין שני מקרים: מורכבות סופית ואינסופית. ופניק הוכיח כי הבעיה ניתנת ללמידה במודל ה PAC אם"ם למחלקה C יש מורכבות סופית[33]. בנוסף הוא הראה כי ככל שהמורכבות של מחלקה נמוכה יותר, הצלחתה להסביר מידע קיים מעידה בהסתברות גבוהה על יכולת ההכללה של ההסבר. זאת באמצעות הוכחה של אי-שוויון דומה לזה שהצגתי בעמוד הקודם:

$$err(h) < e\check{r}r(h,S) + \varepsilon$$

כאשר ε יורד ככל שמדד VC של המחלקה C קטן יותר.

מדד המורכבות של ופניק הינו מושג קומבינטורי מסובך למדי ולא אנסה להראות כאן את הקשר בינו לבין הרעיון של "יכולת הסבר", או של בחינות והפרכה. עם זאת, מן הראוי לציין כי ופניק עצמו הושפע מפופר ובניסוח התיאוריה המתמטית שלו ראה מימוש של רעיונותיו הפילוסופיים של פופר[34].

מאוחר יותר הוכנס לשימוש מדד פורמלי אינטואיטיבי יותר של מורכבות, בשם סיבוכיות רדמכר[35].
לפי מדד זה, מורכבות של מחלקת היפותזות היא החלק היחסי מכל המדגמים האפשריים אותו היא מסוגלת להסביר. גם מדד זה מספק לנו אי-שוויון מהצורה שהראיתי קודם, ובעזרתו ניתן לחסום את השגיאה האמיתית. מעבר לכך, מדד זה ניתן בקלות יחסית להכליל לבעיות מורכבות יותר מבעיות הקלסיפיקציה בהן עסקתי בעבודה זו.

---

[33] Vapnic, *S.L.T.* , p.149, theorem 4.5. הגדרה של ממד VC בנספח ו'.
[34] שם. pp. 106-108
[35] Bartlett & Mendelson (2002). הגדרה פשוטה ואלגנטית של סיבוכיות רדמכר בנספח ז'.

מושגים מקבילים בלמידה ובפילוסופיה

בפרק זה אשווה באופן פרטני יותר בין כמה מושגים בתורה של פופר לבין המושגים המקבילים בלמידה חישובית סטטיסטית.

תצפיות

פופר נוקט גישה אובייקטיבית ודוחה גישות הגורסות כי המתבונן יכול לפרש את אותה התצפית באופנים סותרים, אך עדיין קבילים[36].

הגם שאופן הסתכלות זה מפשט את הבעיות הכרוכות בהפרכה ע"י תצפיות הוא אינו פותר את כולן. עדיין עלינו לבחור **במה לצפות.** לפי פופר עלינו לבצע תצפיות שיהוו מבחנים בלתי תלויים לתיאוריה, וכן שיהיו כמה שיותר קשים ובעייתיים מבחינת התיאוריה, כלומר תצפיות שא-פריורית לא סביר שתוצאותיהן יתיישבו עם התיאוריה החדשה[37]. הדרישה למבחנים "קשים" אינה באה כלל לידי ביטוי במודל הPAC שהצגתי. הסעיף הבא מסביר איזו דרישה מחליפה אותה ומדוע.

האם לכל תיאוריה יש ערך מספרי המתאר את ניתנותה להפרכה?

עקרון התיחום של פופר מאפשר לנו להבחין בין תיאוריות הניתנות להפרכה וכאלה שאינן. פופר אף הציע דרכים להעריך האם תיאוריה מסוימת ניתנת להפרכה במידה רבה או פחותה יותר מרעותה במקרים מסוימים[38]. עם זאת, הוא נמנע מלהקצות ערך מספרי מוחלט שיאפשר לנו להשוות בין כל שתי תיאוריות שהן, בטענה כי מטריקה כזו מחייבת הנחות שהן מחוץ ללוגיקה הבסיסית של תורת ההסתברות עליה הוא מסתמך[39]. הנחות אלה מופיעות בבעיות למידה חישובית בדמות הפילוג D אותו הצגתי קודם, והן תקפות כאשר מלמדים מערכת בה יעשה שימוש בסביבה ספציפית[40]. זה אינו אומר שכדי לחשב את מורכבות מחלקת היפותזות עלינו להכיר את הסביבה בה המערכת עתידה לפעול, מכיוון שהחסם אותו הצגתי קודם **מתקיים עבור כל פילוג**, ובלבד שמובטח לנו שהדוגמאות המשמשות ללמידה יגיעו מאותו הפילוג בו המערכת אמורה לפעול.

פשטות ויכולת הכללה

לאורך העבודה השתמשתי במושג "ניתנות להפרכה" ביחס לתיאוריות מסתכנות עפ"י הגדרתו של פופר, ולעומת זאת במושגים "פשטות" ו"מורכבות" ביחס למחלקות היפותזות. עתה ברצוני להראות כיצד מתקשר מושג הפשטות ליכולת הכללה ולהפרכה, ומכאן לכל תיאוריה מדעית. ציינתי בתחילת העבודה כי פילוסופים ומדענים דרשו מתיאוריה מדעית טובה שתהיה "פשוטה", והשתמשו לשם כך בהגדרות שונות. כל המצדדים בפשטות היו מסכימים כי הפרבולה באיור 1 מהווה הסבר **טוב יותר** מהעקום העובר דרך הנקודות. חלקם היו מצדיקים זאת ע"י

---

[36] Popper, *C&R* p. 63

[37] שם. p.240

[38] למשל כאשר יש סדר הכלה בין קבוצות הפסוקים שיכולים להפריך אותן. Popper, *L.Sc.D* , chapter VI

[39] שם. p. 118, footnote 1

[40] המחשה לכך באמצעות דוגמא בנספח ח'.

שיקולים מתמטיים וסטטיסטיים, חלקם באמצעות טענות על אופיו של הטבע, וחלקם באמצעות טענות על אסתטיות, אורך התיאור הנדרש או שיקולים אחרים. פופר התייחס לגישה זו וטען כי ההגדרה שמשקפת היטב את מושג הפשטות הרצוי במדע מתלכדת עם מושג ה"ניתנות להפרכה" שלו[41]. הוא דחה לחלוטין כל פרשנות אסתטית או תלויית ייצוג של המושג בהקשר המדעי, בטענה שזו חייבת להיות סובייקטיבית וחיצונית ללוגיקה[42].

בשנת 1978 הציג ג'ורמה ריסאנן מדד בשם MDL המגדיר את אורך התיאור המינימלי של אוסף דוגמאות כלשהו[43]. השימושים הרבים וההשלכות הפילוסופיות של MDL זה הינם מחוץ לתחום של עבודה זו, אולם אציין רק כי נובע ממנו שיש קשר הפוך בין אורך התיאור של מחלקת פונקציות לבין מגוון המדגמים אותם היא מסוגלת להסביר. המשמעות היא שככל שניתן לנסח תיאוריה פורמלית באופן קצר יותר, כך קל יותר "להפריך" אותה. האינטואיציה מאחורי מושג הפשטות משקפת למעשה חלק מהעקרונות עליהם מושתתת תורתו של פופר.

---

[41] Popper, *L.Sc.D,* chapter VII

[42] שם. p. 137

[43] Rissanen, *Stochastic complexity and universal modeling* (1994), ראה גם תיאור הקשר ההדוק בין MDL לתער של אוקהם: http://en.wikipedia.org/wiki/Minimum_description_length

ביקורת

חילקתי את הביקורת הקיימת לשלושה סוגים:

- ביקורת על הגישה הפילוסופית של פופר, שטענתי כי הקשר בינה לבין עקרונות הלמידה החישובית הוא הדוק ביותר. אציג כמה טענות הנוגעות למתודת ההפרכה.
- טענות המצביעות על בעיות בתיאוריה של למידה סטטיסטית ועל חולשת הקשר בין תיאוריה זו למערכות פרקטיות שכביכול ממשות אותה.
- טענות כנגד הקשר העקרוני בין פעולה אנושית במהותה כחקירה מדעית, לבין פעולה טכנית-מתמטית של מחשבים.

ביקורת על פופר

- לפי הגישה ההוליסטית, שנציגיה הבולטים הם קווין ודוהם[44]: לעולם לא ניתן להפריך תיאוריה יחידה, מכיוון שהמסקנות נגזרות תמיד מאוסף נרחב של הנחות ותיאוריות. לכן כל שניתן לומר הוא שאם המסקנה הופרכה, אזי לפחות אחת מכל ההנחות או התיאוריות שגויה. במקרה כזה, אם נתייחס למכלול ההנחות והטענות כתיאוריה אחת עלינו לדחות את כולן והמדע לא יצליח להתקדם כלל, כי כולו יקרוס עם כל תצפית לא צפויה. הרעיונות הבסיסיים של פופר אותם הצגתי אינם מספקים לנו דרך לבודד את ההשערה השגויה. פופר מתייחס לגישה זו בספרו L.Sc.D. לא אכנס כאן לפרטי תגובתו.

  בהקשר של למידה סטטיסטית אוסף ההנחות והטענות האפשריות כולו מסתכם באופן בו מתרגמים את הקלט מהעולם האמיתי לייצוגו הממוחשב. בסיטואציות פרקטיות של קלסיפיקציה, אי הצלחה למצוא היפותזה טובה אכן נובעת לרוב מבעיה בתרגום ולא מבעיה בבחירת המחלקה[45].
- לפי פרשנות רלטיביסטית של קון[46], הרעיון שתצפית יכולה לאשש או להפריך מסקנה כלשהי הוא שגוי, מכיוון שהאופן בו המדען מפרש את תוצאות התצפית תלוי בפרדיגמה בה הוא משתמש ובהנחותיו המוקדמות, ואלו עשויות להשתנות בין מדען למדען. יתר על כן, התומכים בגישה מקבלים את האפשרות ששתי פרדיגמות מנוגדות הן אמיתיות, הגם שהן סותרות זו את זו. בפרט לא ניתן להפריך שום תיאוריה, כי כל תצפית המפריכה ניבוי של התיאוריה עפ"י פרדיגמה אחת עשויה להתיישב עם אותה התיאוריה לפי פרדיגמה אחרת, אמיתית באותה מידה. גישה זו לא ניתן ליישב עם מושג האמת שפופר משתמש בו, והוא דוחה אותה על הסף כפי שהראיתי בפרק על אמת מוחלטת ואמת אמפירית.

  פופר מקבל את עצם הרעיון של פרדיגמות ודוגמות מתחרות ואף מודה בחשיבותן להתקדמותו של המדע, אך כל זאת בתנאי שבסופו של דבר אלו יהיו נתונות לחשיבה ביקורתית וניתנות להפרכה, ולא יוכלו לשכון לנצח זו לצד זו[47].

---

[44] http://en.wikipedia.org/wiki/Confirmation_holism

[45] עובדה זו נובעת מאופיין הספציפי של בעיות קלסיפיקציה שימושיות רבות, ראה נספח ט'.

[46] התומכים בגישה זו מסתמכים בין השאר על ספרו של קון "המבנה של מהפכות מדעיות". אף שקון עצמו התנער מפרשנות קיצונית כזו של דבריו באחרית דבר שפרסם מאוחר יותר.

[47] פופר מביע עמדה זו ביחס למיתוסים ודוגמות. "מדע: השערות והפרכות" עמ' 16-17.

ביקורת על תיאוריה של למידה סטטיסטית

גם התורה הפורמלית-מתמטית של מערכות לומדות אינה חסינה לביקורת. אף אחד אינו טוען כמובן שההוכחות של ווֹפניק או ריסאנן פגומות, אך קיים וויכוח לגבי החלות של הטענות התיאורטיות שהצגנו על המציאות.

- כל החסמים שהראיתי מתקיימים בהסתברות בלבד. כלומר תמיד יתכן מדגם "רע" – שעבורו החסם אינו מתקיים וכל המסקנות שהסקנו ממנו אינן מתחייבות (ובפרט המסקנה על יכולת ההכללה – "האמיתיות" – של ההיפותזה שלנו). בהקשר זה אציין רק שמי שאינו מוכן לקבל טענות סטטיסטיות יאלץ לדחות כמעט כל תצפית מדעית עקב אי הדיוק (ולו המזערי) הכרוך בביצוע מדידות כלשהן. עבור מספר דוגמאות מספיק, ההסתברות לאי קיום החסם היא זניחה ממש לעומת אי דיוקים כאלה. הפילוסוף קרל המפל נתן לכך גושפנקה במודל האינדוקטיבי-סטטיסטי. במודל זה להסבר מדעי התיר המפל מסקנות המתקבלות מהתיאוריה באופן זה[48].
- החסם המרכזי שהראיתי משתמש במורכבות מחלקת ההיפותזות ("התיאוריה") כדי לחסום **מלמעלה** את הפער בין השגיאה האמפירית ("היכולת להסביר תצפיות קיימות") לבין השגיאה האמיתית. אין כל הוכחה כי פער זה הוא גדול במקרה של מחלקה מורכבת ("קשה להפרכה"). יתכן שיש דרך אחרת לחסום אותו שאינה משתמשת כלל במורכבות מחלקת ההיפותזות. אם נשחזר טענה זו מהכיוון הפילוסופי, אזי פופר צודק רק בכיוון אחד: אם ניקח תיאוריה מסתכנת היא מבטיחה לנו יכולת הכללה טובה (אם לא תופרך). אך אם ניקח תיאוריה שאינה מסתכנת, ייתכן שיקרה אותו הדבר – פשוט אין לנו דרך להבטיח שכך יקרה (או שלא יקרה).
- ההנחות בבסיס המודלים השונים שהצגנו (למשל דגימות i.i.d) אינן מתקיימות במלואן כמעט אף פעם. זוהי בעיה כללית של מודלים מתמטיים המשמשים אותנו לתאר בעיות בעולם האמיתי. למרבה המזל מתברר שכאשר הדרישות מתמלאות "בקירוב", עדיין יש תועלת במודלים ולכן משתמשים בהם. אין באפשרותי לנסח טענה זו באופן פורמלי וכמובן שלא להוכיח אותה, אך זוהי מתודה מקובלת בפרקטיקה של מדעי המחשב.

למידה כפעולה אנושית במהותה

ניתן לטעון כי עצם ההשוואה בין החקירה המדעית, שהיא פעולה אנושית במהותה, לבין פעולה חישובית טכנית כגון למידה סטטיסטית, מניחה גישה רדוקציוניסטית. כלומר שכל פעולה – שכלית, תפיסתית, מכנית וכו' – ניתנת לתיאור במונחים בסיסיים יותר, ושחשיבה אנושית הינה פשוט אוסף של תגובות כימיות ופיסיקליות. מתסכל למדי לחשוב כי אין הבדל מהותי בין האופן בו אנו לומדים והאופן בו המחשב לומד, ולבסוף ניוותר מאחור ומחשבים יחליפו אותנו גם בלמידה וביצוע מטלות מורכבות כמו שקרה בתחומים אחרים.

השאלה עד כמה ניתן להגן על גישות רדוקציוניסטיות עניינה בעיקר לפילוסופים העוסקים בבעיית גוף ונפש, ולא אתמקד בה. יחד עם זאת, עקרונות משותפים עשויים להימצא במערכות שונות גם אם לא מניחים

[48] המפל, קרל, *פילוסופיה של מדע הטבע*, האוניברסיטה הפתוחה, 1979.

שפעולתן זהה, או אפילו דומה. העקרונות המכאניים והאווירודינמיים הפועלים על כדור גומי וצפרדע הינם זהים, וכדי לקבל זאת אין צורך לטעון שצפרדע היא כדור או ששניהם קופצים באותו אופן.

אוסיף ואציין שכל ההשוואה שהצגתי פה מתייחסת לחלק אחד בלבד של המתודה המדעית – בדיקת השערות. המחשב אינו יכול (בינתיים) להחליט עבור עצמו לאילו פונקציות כדאי לבצע אופטימיזציה, מהי מחלקת היפותזות שכדאי לנסות, וכו'. גם אם נקבל את תהליך הלמידה הסטטיסטית כמקביל לחלוטין לתהליך המדעי של בדיקת השערות, עדיין נותר לאדם חופש מלא בניסוח התיאוריות וההשערות. זאת כפי שפופר לא הציב כל מגבלה על מקור ה"מיתוסים" כפי שכינה אותם.

לסיום, כמה מילות הרגעה. מדענים העובדים כיום על מערכות לומדות מציינים כשאיפה את היכולת להגיע לרמה של ילד בין שנתיים, כך שעדיין נותר לנו זמן מה בטרם ילמדו המחשבים כיצד לשלוט בנו ובעולם...

סיכום

בעבודה זו ניסיתי לבחון כיצד עקרונות מתחום הפילוסופיה של המדע באים לידי ביטוי בתחום הטכני-מתמטי של למידה חישובית סטטיסטית. בפרט ניסיתי להציג את ההגדרות הטכניות של שגיאה אמפירית, שגיאת הכללה ומורכבות, כפורמליזציה של הפילוסופיה של פופר העוסקת בתצפיות, אמיתיות של תיאוריה, ויכולת הפרכה. ראינו כי כדי לקבל היפותזה בעלת יכולת הכללה טובה, יש לבחור מחלקת היפותזות פשוטה ולהסתכן בכך שזו לא תוכל להסביר את אוסף הדוגמאות שבידינו. ממש כשם שפופר דרש מהמדען השערות נועזות, שלא יוכלו להסביר כל תצפית אפשרית. הצגתי מספר מדדים כמותיים המשמשים בלמידה, ביניהן בולטת מורכבות רדמכר. לדעתי זו מביעה בנוסחה מתמטית פשוטה את העקרון המרכזי של פופר כפי שהצגתי אותו בפתיחה – על תיאוריה טובה להסביר את התצפיות הקיימות, וכמה שפחות דברים אחרים.

בבעיות למידה חישובית תמיד עומדת מולנו מטרה ברורה: צמצום שגיאת ההכללה. כל מה שאינו מסייע בידנו להשיג זאת אינו רלוונטי. במהלך העבודה ניסיתי להראות כי ניתן ליישם בלמידה חישובית עקרונות שונים מתחום הפילוסופיה של החקירה המדעית, ויישום זה מאפשר לנו לפתור את הבעיה שלפנינו בצורה טובה יותר. דהיינו, להקטין את שגיאת ההכללה. אני רואה בכך מעין חיזוק לטענה ששימוש בעקרונות אלה בתחומם המקורי (דהיינו החקירה המדעית) הוא מוצדק ומקרב אותנו לאמת. בפרט אני רואה ביכולת לממש את רעיונותיו של פופר חיזוק משמעותי לתורתו, ממש כשם שהניסוי הגרעיני הראשון חיזק את תורת היחסות של אינשטיין, כהדגמה חיה לשימושיות שלה.
בפרק האחרון ניסיתי להראות כי לא רק עקרון ההפרכה של פופר זכה לניסוח פורמלי ולמימוש, כי אם גם רעיונות אחרים ועתיקים יותר, כמו התער של אוקהם.

עם זאת, תהיה זו טעות לבחון רעיונות פילוסופים ולהשוות ביניהם רק על פי מידת השפעתם על למידה חישובית. מאפיינים רבים של החקירה המדעית אינם ניתנים לכימות, או אינם קיימים כלל במסגרת פורמלית כזו. למידה חישובית מתאפשרת תחת הנחות רבות על הסביבה, על התצפיות, דרך ייצוגן והיחסים ביניהן, ועוד. בעיקר היא מחייבת מטריקה מסוימת על הדוגמאות, על היפותזות ועל מידת ההתאמה של תחזיות לתצפיות. כמו כן חלק מהשיטות מניחות את קיומו של איזה "פילוג" מסתורי הנמצא תמיד ברקע ואינו משתנה לעולם. מרביתן מחייבות מספר תצפיות גדול יחסית. נדיר למצוא תחום מדעי בו דרישות אלה מתקיימות במלואן. מה הפילוג של ריאקציות כימיות? איך ניתן להשיג 100 דגימות נוספות של התפתחויות אבולוציוניות? האם תוצאת ניסוי שני הסדקים קרובה יותר לתיאורית הגלים או החלקיקים? שאלות אלה מביאות אותי למחשבה שבמקרה הטוב ההשוואה ללמידה חישובית היא בעלת משמעות רק עבור חלקים ספציפיים של החקירה המדעית.
יתר על כן, הסיטואציות בהן מפעילים כלים חישוביים שונות מהסיטואציה הנפוצה של המדען החוקר: המדען לרוב ניצב מול תצפיות, מנסה להסבירן בעזרת תיאוריות שונות, ואז בודק את השערותיו. מכיוון שבדרך כלל הוא הראשון לנסות תיאוריה זו או אחרת, פעמים רבות הוא יכשל, ינסה תיאוריות שונות, יבצע עוד תצפיות, וחוזר חלילה.

בלמידה חישובית איננו משתמשים בדרך כלל כדי לגלות "אמיתות חדשות". המהנדס העומד בפני בעיית קלסיפיקציה אינו מעוניין ללמוד דבר בעצמו. הוא רוצה ללמד את המערכת. לשם דוגמא, הוא רוצה שהמצלמה בכניסה לבנין תוכל לזהות פרצופים, אף שמצלמות כאלה, או דומות מאד, כבר קיימות ופועלות בבניינים אחרים – אין כאן כל תגלית פורצת דרך.

לשם כך הוא משתמש במחלקות היפותזות שידוע לו שהן שימושיות במקרים דומים[49]. לרוב אין הוא נדרש להמציא מחלקות ממחלקות שונות ולהפריכן בזו אחר זו. עיקר עבודתו היא דווקא החלקים אותם לא הצגתי כאן כלל – תרגום המידע מהעולם האמיתי, תכנון האלגוריתם וכו'.

דומה הדבר למדען המשתמש בתיאוריות קיימות כדי להעריך למשל קבועים פיסיקליים. היכולת לעשות זאת באופן עקבי תלויה בכך שהתיאוריה לא תוכל להסביר כל תצפית אפשרית, כלומר תהיה ניתנת להפרכה. אולם, נדיר למדי שתיאוריות מדעיות מבוססות אכן מופרכות בעקבות מדידות שגרתיות אלה. זוהי הדגמה נוספת לכך שההשוואה ללמידה חישובית מתאימה לחלקים מסוימים של החקירה המדעית יותר מאשר לאחרים.

[49] ראה נספח ט'.

<u>נספחים – הגדרות, משפטים ודוגמאות</u>

נספח א' : The PAC model (Probably Approximatly Correct)

הגדרה:

תהי C מחלקת היפותזות מעל X. C היא **ניתנת ללמידה PAC (PAC learnable)**, אם קיים אלגוריתם L בעל התכונה הבאה:

לכל היפותזת מטרה c in C, לכל פילוג D על X, ולכל $0<\varepsilon<½$, $0<\delta<½$ מתקיים: אם L מקבל כקלט את S (קבוצת דוגמאות בגודל m שנדגמה iid מתוך D, עם התשובה הנכונה c(x) של כל דוגמא), את ε ואת δ , אזי בהסתברות של לפחות δ -1, L יחזיר היפותזה h in C, כך ש $err(h)<\varepsilon$.

(כדי שהלמידה תחשב מעשית, דורשים כי L ירוץ בזמן פולינומיאלי ב $1/\delta, 1/\varepsilon$, ובפרט שמספר הדוגמאות m יהיה חסום בפולינום זה)

המשמעות במילים פשוטות היא שעבור מחלקה הניתנת ללמידה , כמעט כל מדגם (אם הוא גדול מספיק) הוא טוב. כמעט מכל מדגם ניתן לקבל היפותזה שכמעט ואינה טועה – ומכאן השם PAC.

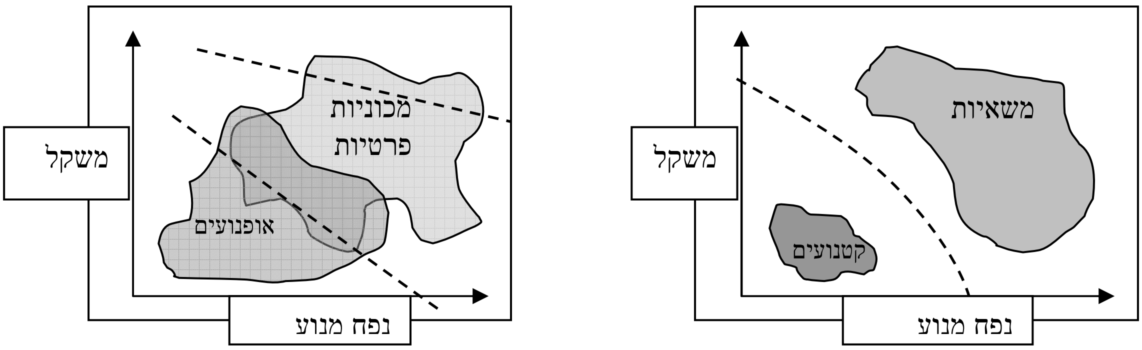
נספח ב': סיווג מושלם וסיווג לא מושלם

כאשר קבוצת הדוגמאות שתשובתן חיובית נפרדת לחלוטין מהדוגמאות שתשובתן שלילית אזי אפשרי סיווג מושלם. דוגמא:

נאמר שעולם הדוגמאות שלנו מורכב מכלי רכב, וכל כלי רכב מיוצג ע"י 2 ערכים – משקל ונפח מנוע.

נרצה היפותיזה – מסווג – אשר תאמר לנו על כל כלי רכב לאיזה סוג הוא שייך.

אמחיש את מושג הסיווג באמצעות איור של "מרחב כלי הרכב". (הנתונים מומצאים לחלוטין)



איור 2.

ניתן לראות שקיים מסווג עבור הבעיה "האם x הוא קטנוע או משאית?" (למשל הקו המקווקו באיור הימני), ואילו לא קיים מסווג כזה עבור הבעייה "האם x הוא אופנוע או מכונית?". ייתכן שקיים מסווג בין מכוניות ואופנועים אם נייצג את כלי הרכב ע"י נתונים אחרים - למשל מספר גלגלים, מהירות מירבית, באמצעות תמונה וכו'.

באמצעות איור זה ניתן גם להמחיש את מושג השגיאה err(h) שהגדרתי. אם נתעלם לרגע מההסתברות השונה להופעת כל פריט (ראה הנספח הבא), אזי השגיאה של היפותזה מיוצגת ע"י השטח הכולל אותו

היא מסווגת לא נכון. החלק היחסי של שטח זה שווה בדיוק להסתברות השגיאה על דוגמא מקרית. כך, עדיין קיימות היפותיזות שהינן "טובות יותר" מאחרות (הקו המקווקו התחתון באיור השמאלי טוב יותר מהקו העליון), מכיוון שהן נותנות לנו תשובה נכונה בהסתברות גבוהה יותר.

נספח ג': כיצד הפילוג משפיע על הערכת ההיפותזות

דוגמא:

נאמר שההיפותיזות h1 ו h2 מסווגות האם רצף של 3 אותיות הוא מילה תקינה בעברית. h1 טועה רק על המילה "בית". h2 טועה על המילים "עלם", "מעש" ו "חלד". שתי ההיפותזות צודקות תמיד על כל שאר המילים. h2 תחשב לטובה יותר, שכן המילים עליה היא טועה נדירות מאוד, ולכן ההסתברות הכוללת שלה לטעות קטנה יותר.

נספח ד': דוגמאות רציפות ומחלקות היפותזות רציפות

לרוב כאשר נרצה לייצג אובייקטים אמיתיים באמצעות וקטור דיגיטלי, עלינו להשתמש במספרים ממשיים, למשל כדי לייצג אורך, משקל, זווית וכו'. אמנם בגלל מגבלות דיוק תמיד ניתן לייצג רק מספר סופי של דוגמאות, אך מספר זה ממילא גדול מכדי שניתן יהיה לבדוק את כל הדוגמאות האפשריות, וההתייחסות אל עולם הדוגמאות כאל עולם רציף מאפשרת פעמים רבות שימוש בכלים מתמטיים המסייעים בביצוע חישובים.

בעולם כזה, כל מחלקת פונקציות עם פרמטרים רציפים (למשל פולינומים) מכילה אינסוף היפותזות שונות, ולכן גודל (קרדינליות) של מחלקת ההיפותזות אינו מדד מספיק.

נספח ה': תיאוריה מסתכנת והסתברות מותנה

נאמר שנתונות 2 תיאוריות כלשהן אותן נרצה לבחון, T1 ו T2.

נסמן ב G1, G2 בהתאמה את האירועים "T1 נכונה" ו "T2 נכונה".

נניח שההסתברות ה א-פריורית של כל אחת מהן להיות נכונה היא p, אזי $p(G1) = p(G2) = p$.

כעת נבצע תצפית O, ונסמן ב E1, E2 בהתאמה את האירועים "T1 מצליחה להסביר את O“ וכנ"ל עבור T2. אם תיאוריה כלשהיא היא נכונה, אזי היא מסבירה את התצפית בהסתברות 1.

כלומר $p(E1|G1) = p(E2|G2)=1$

עתה ניתן לחשב את ההסתברות של כל תיאוריה להיות נכונה בהנתן ששתיהן הצליחו להסביר את התצפית:

$$P(G1|E1) = p(E1|G1)*p(G1)/p(E1) = p/p(E1)$$

$$P(G2|E2) = p(E2|G2)*p(G2)/p(E2) = p/p(E2)$$

אם T1 היא תיאוריה מסתכנת יותר מ T2 , ההסתברות הכוללת שתוכל להסביר את O קטנה יותר:

$$P(E1) < p(E2)$$

ולכן $p(G1|E1) > p(G2|E2)$, כלומר ההסתברות שהתיאוריה T1 אמיתית גבוהה יותר.

נספח ו': מימד VC

מדד קומבינטורי למורכבות של מחלקת היפותזות (בינארית).

נאמר ש VC-dim(C) , הוא מספר הדוגמאות הקטן ביותר הנדרש, כדי שלא ניתן יהיה להסביר אותן בעזרת C. (הגדרה מלאה ודוגמאות בספר עמ' 50-53)

משפט VC:

המחלקה C ניתנת ללמידה PAC אם"ם VC-dim(C) הוא סופי.

עבור מחלקה "עשירה" המסוגלת להסביר הכול (למשל מחלקת "כל ההיפותזות"), VC-dim(C) הוא אינסופי ולכן לא ניתן לקבל היפותזה בעלת יכולת הכללה משום מספר של דוגמאות.

הגרסה הכמותית של המשפט:

תהי C מחלקה ניתנת ללמידה, ונסמן d = VC-dim(C), אזי מספיקות

$1/\varepsilon * \log(1/\delta) + d/\varepsilon * \log(1/\varepsilon)$ דוגמאות כדי שבהסתברות לפחות δ -1 נקבל את היפותזה נכונה (הטועה בהסתברות קטנה מ ε) . כלומר ככל שהמחלקה מסבירה פחות (d = VC-dim קטן יותר), ניתן ללמוד בעזרת פחות דוגמאות. (אם המחלקה מכילה את ההיפותזה הנכונה!)

נספח ז': מורכבות רדמכר (Rademacher Complexity)

הגדרה:

$$Rm(C) = \mathop{\mathrm{E}}_{x}\mathop{\mathrm{E}}_{y}[\sup_{h\in C}\tfrac{2}{m}\left|\sum_{i=1}^{m} y_i \cdot h(x_i)\right|]$$

במבט ראשון אולי קשה לראות את הקשר ההדוק בין נוסחה זו ליכולת הסבר, אך אנסה להבהיר אותו:

X הוא עולם כל הדוגמאות האפשריות, ו Y היא קבוצת התשובות האפשרית ({-1,+1} בבעיות קלסיפיקציה בינארית). נגריל m דוגמאות מקריות לחלוטין, ועבור כל אחת נגריל תשובה מקרית.

הביטוי בערך מוחלט הוא מספר הדוגמאות שהצליחה היפותיזה מסוימת h להסביר, והביטוי בסוגריים הוא החלק היחסי מהמדגם **שההיפותיזה הטובה ביותר** במחלקה הצליחה להסביר.

כעת נעשה זאת עבור כל המדגמים האפשריים (באמצעות לקיחת התוחלת), ונקבל ביטוי המתאר את היכולת של המחלקה C להסביר מדגם אקראי לחלוטין.

כאשר נשתמש במחלקה C כדי למצוא היפותזה שמסבירה את המדגם S (בגודל m), אזי בהסתברות גדולה מ δ -1 מתקיים:

$$|err(h) - e\check{r}r(h,S)| < Rm(C) + A$$

כאשר A הוא קבוע התלוי ב $1/\sqrt{m}$ וב $\ln(\delta)$.

מאי שויון זה נגזור את שני האי שוויונים הבאים:

$$err(h) > e\check{r}r(h,S) - Rm(C) - A$$

$$err(h) < e\check{r}r(h,S) + Rm(C) + A$$

כאשר המורכבות של C קטנה מתקיימים שני הדברים הבאים (בהסתברות גבוהה):

מצד אחד אי-יכולתה להסביר את התצפיות הקיימות ($e\check{r}r(h,S)$ גבוה) משמעה ש C אינה מכילה אף היפותיזה מתאימה ולכן נאלץ לוותר עליה.

מצד שני אם C מצליחה להסביר את התצפיות הקיימות היטב, מובטח לנו שגם שגיאת ההכללה שלה היא קטנה.

כלומר מחלקת היפותזות C שמצליחה להסביר את המידע התצפיתי S, אבל כמה שפחות דברים אחרים היא מוצלחת ובהסתברות גבוהה תאפשר לנו יכולת הכללה טובה.

נספח ח': דגימה מתוך פילוג מבלי צורך לדעת מהו

דוגמא:

נאמר שהמערכת צריכה ללמוד להפריד בין תמונות של גברים ונשים. איזה פילוג מתאר היטב את ההסתברות להופעת כל תמונה ספציפית? לא ניתן לדעת, אך עדיין ניתן לדגום i.i.d מתוך פילוג זה אם ניקח אוסף אקראי של תמונות מאותה מצלמה שאמורה לספק את הקלט למערכת לאחר הפעלתה.

מאחר שההצלחה של המערכת תלויה ביכולתה להפריד בין תמונות שיגיעו מאותה סביבה, כלומר מאותו הפילוג, זה הפילוג היחידי שרלוונטי לפתרון הבעיה.

נספח ט': בעיות קלסיפיקציה קלאסיות

לרוב כאשר מנסים למצוא מפריד בין סוגים שונים של עצמים, הפרטים השייכים לכל אחד מן הסוגים "דומים" זה לזה באופן כלשהו, ו"שונים" מהפרטים השייכים לסוג השני. ניתן לנצל תכונה באמצעות תרגום העצמים לדוגמאות דיגיטליות באופן כזה שישקף את דמיון בין עצמים מאותו סוג. דרך אחת לעשות זאת היא לייצג את הדוגמאות במרחב מטרי (למשל מרחב אוקלידי), בו המרחק בין שתי דוגמאות הוא קטן יותר ככל שהן דומות זו לזו.

אם הצלחנו לעשות זאת היטב אזי כל סוג של עצמים ייוצג ע"י "גוש" כלשהו במרחב זה, ו 2 הגושים לא יהיו קרובים מדי זה לזה (ראה איור 2 מימין). במקרה כזה ישנן מחלקות היפותיזות רבות המכילות מסווג טוב. אם לא הצלחנו אזי אין מסווג טוב ובפרט אף מחלקת היפותיזות לא מכילה מסווג כזה (ראה איור 2 משמאל).

פעמים רבות עיקר המאמץ נדרש לפתרון בעיית הייצוג, ולאחר מכן ניתן להשתמש באחת מכמה מחלקות מוכרות ו"נוחות" מבחינה מתמטית על מנת למצוא מסווג טוב. בין מחלקות אלה ניתן למנות מפרידים ליניאריים, פולינומים, עצי-החלטה ועוד. מדדי המורכבות אותם הצגתי הם נמוכים למדי עבור מחלקות אלה, ועם זאת הן משמשות בהצלחה במגוון רחב של בעיות פרקטיות.

ביבליוגרפיה

אריסטו. *המטאפיסיקה, ספר א'*. תשנ"ח. מיוונית: ח"י רות.

המפל, קרל ג. *פילוסופיה של מדע הטבע*, האוניברסיטה הפתוחה, 1979.

יום, דיוויד. *מסכת טבע האדם*, ספר 1, הוצאת מאגנס, תשי"ד. מאנגלית: יוסף אור.

פופר, קרל ר. *מדע: השערות והפרכות* (מאמר מתורגם בספר מקורות האוניברסיטה הפתוחה). מאנגלית: יורם נבון.

קון, תומס ס. *המבנה של מהפיכות מדעיות*. תל אביב (1977). מאנגלית: יהודה מלצר.

Bartlett, Peter L. and Mendelson, Shahar. *Rademacher and Gaussian Complexities: Risk Bounds and Structural Results*. Journal of Machine Learning Research Vol. 3 (2002), pages 463-482.

Burney, S. M. A., Jilani ,T. A. and Ardil, C. *Levenberg-Marquardt Algorithm for Karachi Stock Exchange Share Rates Forecasting*. Enformatika, Volume 3, pp. 316-320

Goodman, Nelson. *The new riddle of induction.* appeard in Fact, Fiction, and Forecast *(1983).*

Kearns, Michael J. and Vazirani, Umesh V. *An Introduction to Computational Learning Theory.* The MIT press (1994)

Popper, Karl R. *Conjectures and Refutations.* Routledge and Kegal Paul, 4th edition (1973).

Popper, Karl R. *The Logic of scientific discoveries.* Hutchinson (1959).

Rissanen, Jorma. *Stochastic complexity and universal modeling.* A lecture in IEEE International Symposium on Information Theory (1994)

Vapnic, Vladimir N. *Statistical Learning Theory* , Wiley-Interscience (1998)

Articles from the web:

Empiricism - http://plato.stanford.edu/entries/rationalism-empiricism/ *viewed on 1/12/06*

Lagrange polynomial -
http://mathworld.wolfram.com/LagrangeInterpolatingPolynomial.html *Viewed on 26/10/06*

MDL - http://en.wikipedia.org/wiki/Minimum_description_length *Viewed on 26/10/06*

Occam's razor & simplicity - http://plato.stanford.edu/entries/simplicity/ *viewed on 29/11/06*

The Quine-Duhem thesis - http://en.wikipedia.org/wiki/Confirmation_holism *viewed on 29/11/06*

בנוסף חלק מן הרעיונות נסמכים על ההרצאות בקורסים אשנב לפילוסופיה של המדע (פרופ' ימימה בן מנחם), IML (פרופ' אמנון שעשוע), תורת האינפורמציה (פרופ' תלי תשבי) ולמידה סטטיסטית (מר עופר דקל). כל השגיאות הן שלי בלבד.