

מבוא ללמידה חישובית | סיכום שיעור 1 (20942)

מתן כהן

סוגי למידה

למידה מונחית/למידה מפקחת - Supervised Learning

- מתקבלת קבוצת אימון
- הדוגמאות בקבוצת האימון הן **מידע מתויג** שהוא זוגות הכוללות: <קלט (מאפיינים), פלט>
- כמו-כן מתקבלת קבוצה של דוגמאות המבחן (Test data)
- הדוגמאות הן דוגמאות שלא נצפו בעבר אבל אין להן את התיוג

למידה לא מפקחת - Unsupervised Learning

- מתקבלת קבוצה לא מתויגת של נתונים
- הסיבות:
- (1) יכול להיות שתהליך התיוג מאוד יקר
- (2) אין דרך לתייג את המידע בצורה טובה
- שימוש ב-Clustering
- חלוקת המידע לקבוצות כך שבכל קבוצה האובייקטים דומים אחד לשני ושונים מאובייקטים בקבוצות האחרות

למידת מחזוקים - Reinforcement Learning

- ישנו סוכן שמתקשר עם הסביבה והמטרה שלו היא לקבל החלטות על מנת למקסם פונקציית מטרות כלשהי בעתיד (Reward)
- בכל שלב הסוכן יכול להשתמש במידע שקיבל עד כה על מנת לבצע החלטה או לחלופין לנסות גישה חדשה כלשהי על מנת להמשיך בתקשורת עם הסביבה
- הפוקוס כאן הוא על ביצועים - צריך למצוא איזון בין ניצול (exploitation) לחקירת הסביבה (exploration), ניסוי של אלטרנטיבות חדשות כאלו ואחרות אל מול שימוש בסט פעולות שכבר מוכרות לסוכן
- התוצאה שהסוכן מקבל לפונ'ציית המטרה מתקבלת בסוף כל שלב (ממש כמו במשחק מחשב)

למידה אקטיבית אל מול למידה פסיבית

- למידה אקטיבית היא מצב בו הלומד מתקשר עם הסביבה תוך כדי הלמידה ו**מבקש** לעשות ניסויים תוך כדי הלמידה או **מבקש לתייג** דוגמאות ספציפיות **במהלך הלמידה** - הכל בזמן האימון
- למידה פסיבית היא מצב הפוך - **הלומד "מסתכל מהצד"** ולומד ולאחר מכן מקבל החלטות - אין לו שום השפעה על תהליך הלמידה וקבוצת האימון

למידה מקוונת (Online) אל מול למידה ממקבץ (Batch)

- לימוד מקוון - Online Learning: הלומד מנפיק תיוגים תוך כדי למידה ומקבל על כך פידבק בזמן אמת
- לימוד ממקבץ - Batch Learning: הלומד מקבל את כל המידע מראש, מעבד את כולו, לומד אותו ולאחר מכן מנפיק ניבויים על סמך המידע

למידה מפקחת - Supervised Learning

הגדרה

- מתקבלות זוגות של $\langle x, y \rangle$ פלט, קלט
- ישנה פונקציה (פונקציית המטרה) **שלא ידועה לנו**: $f : X \rightarrow Y$ כך ש:
 - ★ X הוא מרחב הקלט (סט כל הקלטים האפשריים) x כך שכל x הוא ממרחב d -מימדי (\mathbb{R}^d) (נניח \mathbb{R}^d)
 - הערה: מעתה נרשום את הוקטור x ללא קו תחתון, נזכור שהוא x קטן
 - ★ Y הוא מרחב הפלט (סט כל הפלטים האפשריים) - בד"כ סקלר
 - במידה והפלטים ב- Y הם בדידים (דיסקרטים) - נקרא לבעיה **בעיית סיווג** (קלאסיפיקציה)
 - במידה והפלטים ב- Y הם רציפים - נקרא לבעיה **בעיית רגרסיה**
- סט כל הצמידים $\mathcal{D} = \{(x_i, y_i) \mid x_i \in X, y_i = f(x_i) \ i = (1, \dots, N)\}$ הוא **סט הנתונים** (Data set)
- אלגוריתם הלמידה \mathcal{A} מקבל את סט הנתונים \mathcal{D}
- נניח שקיים פילוג כלשהו $P(x)$ שמייצר דוגמאות **אימון** ודוגמאות **מבחן** באותה התפלגות
 - ★ כיוון שרק כך נוכל ללמוד מדוגמאות
- נגדיר \mathcal{H} להיות **סט ההיפותזות** שהוא מכלול הפונקציות שאנו בוחרים שהאלגוריתם יכול להשתמש בהן על מנת לחזות את הדוגמאות
 - ★ מתוך כלל הפונקציות נבחר את הטובה ביותר אשר **לדעתנו** מתארת בצורה הטובה ביותר את הקשר בין ה- x ים לבין ה- y ים בקבוצת האימון בהנחה שהקשר יישמר (על פי אותה הנחה על P)
 - ★ במילים אחרות: **נרצה לבחור מועמדת g עבורה**: $g(x) \approx f(x)$ ועל מנת לבצע זאת:
 - נרצה לייצג מדד ביצועים לכל אחת מהפונקציות ב- \mathcal{H} על קבוצת האימון
- נעשה זאת על ידי שימוש בפונקציה **in-sample error** ($E_{in}(h)$) שתייצג את השגיאה של כל $h_i \in \mathcal{H}$:

$$\forall h_i \in \mathcal{H} : E_{in}(h_i) = \frac{1}{N} \sum_{n=1}^N [(h_i(x_n) \neq f(x_n))]$$

◁ בעזרת פונקציה זו נוכל לנפות פונקציות לא טובות מתוך סט ההיפותזות

שגיאות - Error Measures

- בכל הקשור לזיהוי שגיאות ובהקשר למדד הביצוע של ההיפותזות מסט ההיפותזות שלנו, יש צורך להבחין כיצד נרצה לסווג בין אלגוריתמים
- יש צורך להבחין בין מדדי שגיאה שונים לבעיות שונות ולבחון מה חשוב לנו:
 - ★ האם אנחנו מעדיפים הרבה **שגיאות קטנות** על פני **מעט שגיאות גדולות**?
 - ★ האם אנחנו מעדיפים מצבים של **false alarm** על פני מצבים של **miss detection**?

המודל הלינארי

- משפחת המודלים הלינארים היא משפחה מאוד שימושית באלגוריתמי למידה מכמה סיבות בסיסיות:
 - ★ חישובים יעילים
 - ★ מתאים לדוגמאות מהטבע - נותן הצדקה לשימוש
 - ★ מודלים אינטואיטיביים שקל מאוד לפרש

מבנה המודל הלינארי הבסיסי

$$\left(\sum_{i=1}^d w_i x_i \right) + b$$

- x_i - קלט ממרחב d -ממדי

- w_i - סקלר - משקל המוכפל בקלט

- b - סקלר קבוע (במודל אפיני)

נרצה להרכיב פונקציות שונות על הפונקציה הלינארית ובכך נקבל חזאים שונים ולכן גם טיפול בבעיות שונות.
לדוגמה:

- הרכבה של פונקציה שהטווח שלה $\{0, 1\}$ תייצר מודל של Binary Classification

- הרכבה של פונקציה שטווח הערכים שלה הוא ממשי תייצר מודל של Linear Regression

- הרכבה של פונקציה שטווח הערכים שלה הוא $[0, 1]$ תייצר מודל של Logistic Regression