

פרויקט מעשי במבוא לבינה מלאכותית תשפ"ד

מרצה: ראובן כהן מתרגלת: יעל משה

ניתן (ורצוי) לעבוד בזוגות. מספיק שאחד מחברי הקבוצה יגיש את העבודה במודל. כתבו בבירור את שמות המגישים בראש העבודה.

בחרו מסד נתונים כלשהו מהאתר <https://www.kaggle.com/datasets> (אם אתם מעדיפים להשתמש בנתונים ממקור אחר, אנא פנו אלי לאישור). בעזרת השיטות שלמדנו ושיטות נוספות לפי בחירתכם, בנו אלגוריתם למידה מתאים שיסווג או יאשכל את המידע. עליכם להעלות קובץ zip המכיל את הקוד שכתבתם וכן דו"ח מסכם המתייחס לנקודות הבאות:

1. קישור למסד הנתונים בו השתמשתם, וכן תיאור של הנתונים (סוג הנתונים, כמות הדגימות, באלו פיצ'רים השתמשתם מתוך הנתונים וכו').
 2. תיאור האלגוריתמים ושיטות הלמידה בהן השתמשתם, וכן הסבר מדוע בחרתם בשיטות אלה ומדוע הן מתאימות לסוג בנתונים וסוג הניתוח הרצוי.
 3. תיאור של פונקציית הסיכון בה השתמשתם, הדרך בה הערכתם את הביצועים, וכיצד בחרתם הפרמטרים של האלגוריתם בהתאמה למטרות.
 4. ניתוח של התוצאות מבחינת איכות הסיווג, זמן ריצה (והחומרה בה השתמשתם), מספר איטרציות, AUC, קצב התכנסות וכדומה. כולל גרפים לפי הצורך.
 5. הסבירו כיצד חילקתם את הדאטה לקבוצות training, validation, test. שימו לב שלא לערבב בין הקבוצות השונות.
 6. הסבירו את בעיית over-fitting. הדגימו כיצד היא מתקבלת אצלכם בדאטה (אם מתקבלת). ממשו את הדרכים שלמדנו כדי למנוע זאת (dropout, early stopping, regularization). ציינו מה ההבדלים בין שיטות הרגולריזציה השונות, נסו שילובים שלהן ודווחו מה עבד הכי טוב.
 7. מצאו את ההיפר-פרמטרים הכי טובים. הסבירו כיצד מצאתם.
- בכל סעיף התאימו את הסקירה לאלגוריתמים שבחרתם ולסוג הנתונים. הציון ייקבע לפי איכות התוצאות שקיבלתם וכן שימוש נכון באלגוריתמים והתאמתם לסוג הנתונים, איכות הדו"ח, שילוב טכניקות ומקוריות.
- הערה: ניתן לבצע השוואה בין שיטות שונות. מומלץ להשתמש בשיטות הכלליות וה"כבדות" ובעיקר ברשתות נוירונים תוך שימוש בחבילות שלמדתם בתרגיל, הן כדי לרכוש ניסיון בעבודה עם שיטות אלה, והן כדי לקבל תוצאות איכותיות.

בהצלחה!