

כריית נתונים ב-R // תרגיל בית מספר 4

הגשה 11/6

חלק ראשון: עצי רגרסיה (6 נקודות)

בתרגיל זה נשתמש במסד הנתונים של חברת הייטקס מתרגיל 3. מטרתנו תהיה להתאים מודל של עץ החלטה למסד זה. שימו לב כי עץ ההחלטה פה הינו מבוסס רגרסיה ולא סיווג! על כן הוא נקרא עץ רגרסיה (regression tree)

1. (3 נק') לדעתכם, ולפני הרצת המודלים, עבוד מסד זה האם נכון יותר להריץ רגרסיה לינארית או עץ רגרסיה? **הסבירו את תשובתכם בקצרה.**
2. (2 נק') התאימו עץ רגרסיה ל- training data. השתמש בפונקציית ctree. **הציגו את העץ שנוצר.**
3. (2 נק') בצעו טרנספורמציות לוג כנדרש בתרגיל בית 3, ושנו את סוג המשתנים הקטגוריאליים למשתנים מסוג factor. התאימו עץ חדש, **והציגו אותו.**
4. (3 נק') בדקו את ביצועי העצים על ה- validation. **דווחו את מדדי ה-RMSE, והשוו אותם לאלו שמצאתם בתרגיל בית 3.**

אופן הגשה:

- ✓ הגשה דרך אתר למידה
- ✓ יש להגיש קובץ R מתועד.
- ✓ תשובות מילוליות להגיש **בקובץ פלט.**