

תרגיל מעשי, חלק ב'

סטטיסטיקה

1. חשבי את העוצמה למציאת שינוי של סטיית תקן אחת בממוצע ערכי הגלוקוז בין חולי סכרת לשאר המטופלים.
2. בצעי מבחן t לבדיקת ההשערה כי ישנו הבדל בין ערכי הסוכר בדם הממוצעים בין חולי סוכרת לשאינם חולים.
3. בצעי רגרסיה לינארית באמצעות statsmodels לבחינת ההשערה כי ישנו קשר בין bmi בבדיקה הראשונה לערכי הכולסטרול.
4. בדקי השערה הזו שנית, הפעם תוך התחשבות במשתנה המין בנוסף.

למידת מכונה

1. בחרי משתנה מטרה ואסטרטגיה לעיבוד הדאטא לחיזוי. חישבי על סוג משתנה המטרה (רגרסיה מול קלסיפיקציה), התפלגות וגודל המדגם הקיים. איך תוכלי להתאים את הדאטא כך שיהיה שימושי ע"י המודל? נסי לחשוב מה עשויה להיות המשמעות של פרדיקציה כזו בקונטקסט הקליני בתרגילים הבאים נעבוד על בעיית רגרסיה - אמידת לחץ דם סיסטולי.
2. בני מודל dummy, איזה מודלים כאלו עשויים להתאים? הציגי 2 ומדדי את המסנן שלהם.
3. בני מודל רגרסיה פשוט, בחרי 2-3 משתנים שעשויים לחזות טוב (לא לחץ דם דיאסטולי באותה מדידה) ובני מודל רגרסיה לינארי על פיהם (statsmodels או sklearn).
4. נסי לעצב משתנים יצירתיים, האם הם משפרים את הפרדיקציה?
5. בני מודל עם כלל המשתנים שעובדת, נסי מודלים מסוגים שונים ועקבי את המסנן הטוב ביותר שהשגת.