

דף תרגילים עשירי ביסודות מדעי המחשב

לדף התרגילים השבוע מצורף קובץ טקסט ובו מידע על מאות תלמידים, הכולל את הציון הסופי שלהם במתימטיקה, בנוסף לעשרות נתוני תכונות אחרים. המידע נלקח ממאגר פתוח בקישור

<http://archive.ics.uci.edu/ml/datasets/Student+Performance>

יחד עם קובץ הנתונים מצורף קובץ קצר עם סיכום התכונות. המשימה שלכם היא למצוא קבוצת תכונות הרלוונטיות למודל המבקש לחזות סיכויי הצלחה במתימטיקה עבור סטודנט עם נתונים מסויימים. בדף תרגילים זה נלך בצעדים על מנת להגיע לפיתרון.

1. ראשית, עלינו לתת גישה לקובץ הטקסט מתוך העורך. שימרו את הקובץ באותה תיקייה בה אתה שומרים את קובצי הפייטון שלכם. בעורך (Thonny) רשמו:
`f=open("Data.txt")`
מצד שמאל זהו שם משתנה לבחירתכם ובתוך הסוגריים שם שנתתם לקובץ הטקסט. המשתנה לתוכו פתחתם את הקובץ הוא למעשה אוסף של מחרוזות. כל מחרוזת היא שורה בקובץ הטקסט, כשבין ערכי התכונות יש את הסימן ";". עליכם ליצור שתי רשימות של ערכי תכונות, אחת עבור תכונה 16 והשנייה עבור תכונה 17. כלומר, רשימה אחת עם ערכי תכונה 16 עבור כל הסטודנטים, ורשימה שנייה עבור תכונה 17. תכונות אילו הן בינאריות. הרשימות צריכות להיות עם מספרים ולא מילים, לכן התאימו 1 ל "yes" ו 0 ל "no".
2. כתבו פונקציה המחשבת את המרחק האוקלידי בין שני וקטורים j, i דו מימדיים (שתי תכונות).
3. כתבו קוד ליצירת שלוש רשימות. בהתייחס לתכונות 16, 17, ברשימה הראשונה צריכים להיות כל המרחקים האוקלידיים בין כל זוגות הסטודנטים שציונם הסופי הוא יותר מ 10, ברשימה השנייה צריכים להיות המרחקים האוקלידיים בין כל זוגות הסטודנטים שציונם הסופי 10 או פחות, וברשימה השלישית צריכים להיות המרחקים בין כל זוגות הסטודנטים שאחד קיבל יותר מ 10 והשני 10 או פחות.
4. כתבו קוד לחישוב הממוצע של כל רשימה והדפסתו. ציינו לעצמכם אם ומה ניתן ללמוד מהתוצאות.
5. עתה חזרו על תרגילים 1-4 עבור תכונות 18, 21. ציינו לעצמכם אם יש הבדלים ניכרים בין תוצאות אילו לקודמות.
6. כיתבו קוד דומה עבור תכונות 14, 25. שימו לב שעכשיו התכונות הן לא בינאריות, ויכולות לקבל יותר משני ערכים שונים. כדי להשתמש בתכונות אילו, נרמלו את ערכיהן כך שהמקסימלי יהיה 1 והמינימלי 0.
7. בחרו 2 תכונות (או יותר) שלדעתכם רלוונטיות להצלחה בלימודי המתמטיקה, וחזרו על האלגוריתם המתואר בתרגילים קודמים עם תכונות אילו.

בהצלחה!