

به نام خدا

دانشگاه صنعتی امیرکبیر

پروژه نرم افزار درس تحلیل داده و اطلاعات

1. Given 1 - D data set $X = \{-5.0, 23.0, 17.6, 7.23, 1.11\}$, normalize the data set using:
 - (a) decimal scaling on interval $[-1, 1]$,
 - (b) min-max normalization on interval $[0, 1]$,
 - (c) standard deviation normalization.
2. The number of children for different patients in a database is given with a vector $C = \{3\ 1\ 0\ 2\ 7\ 3\ 6\ 4\ 2\ 0\ 0\ 10\ 15\ 6\}$:
Find the outliers in set C using standard statistical parameters mean and variance.
If the threshold value is changed from ± 3 standard deviations to ± 2 standard deviations, what additional outliers are found?

3- الگوریتم مربوط به روش خوشه بندی K-means، را در نرم افزار Python یا R کد نویسی نموده و بر روی دادگان Iris پیاده سازی و نتایج به دست آمده را گزارش نمایید. در مورد دادگان Iris توضیح دهید (توجه: از تابع K-means و کتابخانه SCIKIT LEARN استفاده نکنید؛ در ضمن داده های Iris را میتوانید از <https://archive.ics.uci.edu/ml/datasets/iris> دانلود نمایید.)