

پروژه عمومی

● بخش اول

با استفاده از یکی از کتابخانه‌های کار با داده در پایتون (برای مثال pandas)، [فایل اکسل](#) را ورودی بگیرید.

● بخش دوم *

هدف از این بخش آشنایی بیشتر شما با مجموعه داده‌های ورودی است.

- ۱۰ سطر اول داده‌ها را چاپ کنید.
- تعداد سطرها و ستون‌ها را چاپ کنید.
- به ازای هر ویژگی (ستون) مقدار کمینه و بیشینه‌ی آن را چاپ کنید.

● بخش سوم

میانگین و انحراف معیار (standard deviation) هر ویژگی (ستون) را بدست آورید و داده‌های آن ستون را طوری نرمال‌سازی کنید که میانگین برابر صفر و انحراف معیار برابر یک شود.

● بخش چهارم

داده جدید را به همراه ویژگی‌هایش ورودی گرفته و هر ویژگی آن را با استفاده از میانگین و انحراف معیار آن ویژگی در داده‌های اولیه نرمال‌سازی کنید.

● بخش پنجم

هر ویژگی از داده‌ی جدید را با همان ویژگی از داده‌های اولیه مقایسه کرده و اختلاف (فاصله‌ی) آن‌ها را حساب کنید (با استفاده از متر اقلیدسی euclidean distance) و اختلاف‌ها را مرتب (sort) کنید و k داده‌ای که کمترین اختلاف را با داده‌ی جدید دارند انتخاب کنید. با استفاده از رای‌گیری بین این k داده‌ی انتخاب شده، کلاس داده‌ی جدید را بدست آورید و خروجی بدهید.

● بخش ششم *

با استفاده از کتابخانه‌ی matplotlib به ازای هر ویژگی نمودار میله‌ای ده داده را به صورت رندوم قبل و بعد از نرمال‌سازی نمایش دهید. می‌توانید برای آشنایی و یادگیری matplotlib از این دو منبع استفاده کنید:

<https://matplotlib.org/>

https://www.w3schools.com/python/matplotlib_pyplot.asp

Cheat Sheets:

Pandas: https://pandas.pydata.org/Pandas_Cheat_Sheet.pdf

Matplotlib: <https://matplotlib.org/cheatsheets/handout-beginner.pdf>

NumPy: <https://intellipaat.com/blog/tutorial/python-tutorial/numpy-cheat-sheet/>

بخش‌های ستاره دار اختیاری (optional) هستند و نمره‌ی امتیازی به آن‌ها تعلق می‌گیرد.

