

داده‌های مربوط به اندازه‌گیری متغیرهای مختلف کیفیت آب در نمونه‌های برداشت‌شده از تعدادی ایستگاه هیدرومتری در فایل‌های پیوست با پسوند csv، ارائه شده‌اند. فایل داده‌هایی که توسط هر یک از دانشجویان باید مورد استفاده قرار گیرد با شماره‌ی دانشجویی وی مشخص شده است. ستون متغیر رسته‌ای class در گام‌های (الف) تا (ه) تمرین حاضر کاربردی ندارد و نباید در این گام‌ها اثرگذار باشد.

گام‌های این تمرین به شرح زیر هستند:

الف) بر اساس مقدار ضریب همبستگی، از بین ویژگی‌ها یا متغیرهای دارای همبستگی بالا، یکی را انتخاب کنید و سپس با استفاده از روش استانداردسازی، داده‌ها را پیش‌پردازش کنید. (۲ امتیاز)

ب) خوشه‌بندی سلسله‌مراتبی را روی داده‌های استاندارد شده اجرا کنید. (۵ امتیاز)

در اجرای خوشه‌بندی سلسله‌مراتبی:

۱. از ستون آخر داده‌ها که مربوط به متغیر دبی جریان است در خوشه‌بندی استفاده نکنید. (دبی جریان جزء مؤلفه‌های بردار ویژگی‌های مورد استفاده در خوشه‌بندی نباشد).

۲. برای سنجش فواصل نقاط از رابطه‌ی فاصله‌ی اقلیدسی استفاده کنید.

۳. از بین روش‌های اجرای خوشه‌بندی سلسله‌مراتبی single linkage، complete linkage، average linkage و Ward، حداقل یک روش را برای اجرای خوشه‌بندی انتخاب کنید. استفاده از بیش از یک روش در صورت ارائه و تحلیل مناسب نتایج، دارای امتیاز اضافه خواهد بود.

۴. خوشه‌بندی‌ها را به‌ازای تعداد خوشه‌ها از ۲ تا ۵ اجرا کنید. اجرای خوشه‌بندی با تعداد خوشه‌های بیشتر، در صورت ارائه و بررسی نتایج، دارای امتیاز اضافه خواهد بود.

۵. نتیجه‌ی اصلی که در مورد هر خوشه‌بندی باید ارائه شود، چگونگی اختصاص نقطه‌داده‌ها یا ایستگاه‌ها به خوشه‌ها است. (یعنی مشخص شود در هر خوشه‌بندی کدام ایستگاه‌ها به عضویت کدام خوشه‌ها درآمده‌اند).

ج) خوشه‌بندی افرازی را با استفاده از الگوریتم K-means روی داده‌های استاندارد شده اجرا کنید. (۴ امتیاز)

در اجرای خوشه‌بندی افرازی:

۱. از ستون آخر داده‌ها که مربوط به متغیر دبی جریان است در خوشه‌بندی استفاده نکنید. (دبی

جریان جزء مؤلفه‌های بردار ویژگی‌های مورد استفاده در خوشه‌بندی نباشد).

۲. برای سنجش فواصل نقاط از رابطه‌ی فاصله‌ی اقلیدسی استفاده کنید.

۳. تعیین مقدار آرگومان‌های تابع اجرای الگوریتم مانند چگونگی تعیین مراکز اولیه و تعداد تکرار

محاسبات الگوریتم و توجیه دلیل انتخاب آن‌ها بر عهده‌ی شما است.

۴. خوشه‌بندی‌ها را به‌ازای تعداد خوشه‌ها از ۲ تا ۵ اجرا کنید. اجرای خوشه‌بندی با تعداد خوشه‌های

بیشتر، در صورت ارائه و بررسی نتایج، دارای امتیاز اضافه خواهد بود.

۵. نتیجه‌ی اصلی که در مورد هر خوشه‌بندی باید ارائه شود، چگونگی اختصاص نقطه‌داده‌ها یا

ایستگاه‌ها به خوشه‌ها است. (یعنی مشخص شود در هر خوشه‌بندی کدام ایستگاه‌ها به‌عضویت

کدام خوشه‌ها درآمده‌اند).

د) شاخص صحت خوشه‌ی میانگین silhouette width را برای تمام خوشه‌بندی‌های اجرا شده محاسبه کنید

و بر اساس آن کیفیت خوشه‌بندی‌های اجرا شده را با هم مقایسه کنید. (۲ امتیاز)

ه) خوشه‌بندی فازی را با استفاده از الگوریتم c-means روی داده‌های استاندارد شده اجرا کنید. (۵ امتیاز)

در اجرای خوشه‌بندی فازی:

۱. از ستون آخر داده‌ها که مربوط به متغیر دبی جریان است در خوشه‌بندی استفاده نکنید. (دبی

جریان جزء مؤلفه‌های بردار ویژگی‌های مورد استفاده در خوشه‌بندی نباشد).

۲. برای سنجش فواصل نقاط از رابطه‌ی فاصله‌ی اقلیدسی استفاده کنید.

۳. تعیین مقدار آرگومان‌های تابع اجرای الگوریتم مانند چگونگی تعیین مراکز اولیه و تعداد تکرار

محاسبات الگوریتم و توجیه دلیل انتخاب آن‌ها بر عهده‌ی شما است.

۴. خوشه‌بندی‌ها را به‌ازای تعداد خوشه‌ها از ۲ تا ۵ اجرا کنید. اجرای خوشه‌بندی با تعداد خوشه‌های

بیشتر، در صورت ارائه و بررسی نتایج، دارای امتیاز اضافه خواهد بود.

۵. نتیجه‌ی اصلی که در مورد هر خوشه‌بندی باید ارائه شود، چگونگی اختصاص نقطه‌داده‌ها یا

ایستگاه‌ها به خوشه‌ها است. (یعنی مشخص شود در هر خوشه‌بندی کدام ایستگاه‌ها به‌عضویت

کدام خوشه‌ها درآمده‌اند).

و) شاخص‌های صحت خوشه‌ی فازی (ضریب افراز، انتروپی افراز، شاخص فوکویاما - سوگنو، شاخص ژی -

بنی) را برای تمام خوشه‌بندی‌های فازی اجراشده محاسبه کنید و بر اساس آن کیفیت خوشه‌بندی‌های اجراشده

را با هم مقایسه کنید. (۲ امتیاز)

نکات مربوط به نحوه‌ی ارائه‌ی پاسخ تمرین

- پاسخ تمرین شامل فایل (یا فایل‌های) برنامه‌ی (یا برنامه‌های) نوشته‌شده به زبان R (۱۲ نمره شامل

۱۰ نمره برای صحت کدها و نتایج و ۲ نمره برای کیفیت کدنویسی) و یک گزارش متنی (فایل با

قالب doc یا docx). حداکثر در ۱۰ صفحه (۸ نمره شامل ۴ نمره برای صحت نتایج و ۴ نمره برای

کیفیت گزارش) است. شایان ذکر است که هیچ بخشی به‌تنهایی دارای امتیاز نیست.

- گزارش متنی، غیر از صفحه‌ی عناوین و مشخصات باید حداقل شامل سه بخش بیان مسأله، مواد و

روش‌ها (شامل معرفی داده‌ها و روش حل مسأله) و نتایج و بحث (شامل ارائه‌ی نتایج و بحث و تحلیل

روی آن‌ها) باشد.

- سعی کنید برای ارائه‌ی نتایج و تحلیل آن‌ها حتی‌الامکان از نمودارها و جداول استفاده کنید.

- برای ترسیم نمودارها از زبان و محیط نرم‌افزاری R استفاده کنید.



به نام خدا

نام درس: «مدیریت کیفی منابع آب»
نیمسال اول سال تحصیلی ۱۴۰۴-۰۵
تمرین پایان ترم
تاریخ ارسال: ۱۴۰۴/۱۰/۱۵
موعد تحویل پاسخ: ۱۴۰۴/۱۰/۳۰

- مجموعه‌ی فایل‌های مذکور را در قالب یک پوشه‌ی فشرده با قالب zip یا rar از طریق سامانه‌ی مدیریت یادگیری الکترونیک دانشگاه (<https://lms.sbu.ac.ir/>) ارسال کنید.

سربلند و پیروز باشید!

آهنی