

به نام ایزد منان



دانشکده مهندسی کامپیوتر



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

تمرین سری سوم داده کاوی

توضیحات:

- پاسخ به تمرین‌ها باید به صورت انفرادی صورت گیرد. حداقل برخورد با پاسخ‌های مشابه، تخصیص نمره کامل منفی به طرفین خواهد بود.
- پاسخ‌های خود را به زبان فارسی و به صورت مرتب، در قالب یک فایل فشرده (.zip) با الگوی زیر در صفحه‌ی درس بارگذاری کنید:

DM_HW[No]_[Student_number].pdf
- لطفاً نظم، ساختار و توالی سوالات را در پاسخ‌ها رعایت کنید.
- تمیزی و خوانایی گزارش تمرین از اهمیت بالایی برخوردار است.
- برای تمرین‌های عملی، علاوه بر کد گزارش کتبی نیز ارسال کنید.
- در صورتی که درمورد این تمرین سوال یا ابهامی داشتید با ایمیل dm.1401.spring@gmail.com با تدریس‌یاران درس در ارتباط باشید.
- مهلت ارسال تمرین تا ساعت ۱۱:۵۵ روز جمعه مورخ ۲۰ خرداد ۱۴۰۱ است.

نیم‌سال دوم ۱۴۰۰-۱۴۰۱

صفحه

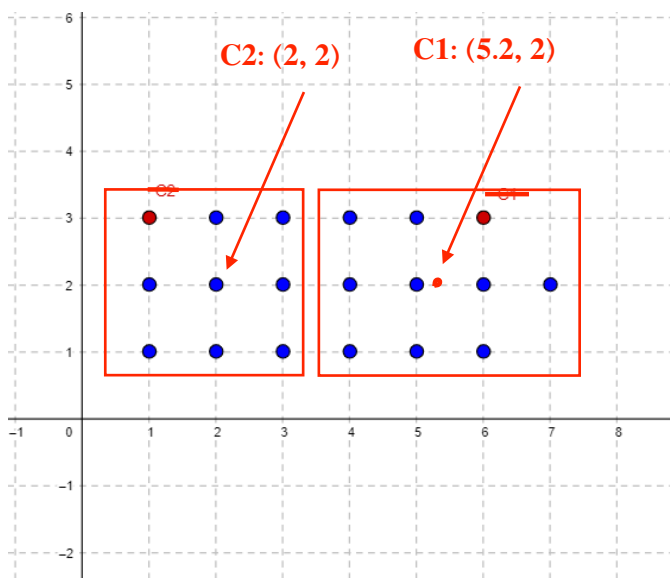
فهرست مطالب

۳	بخش نوشتاری.....
۳	سوال اول.....
۳	سوال دوم.....
۴	سوال سوم.....
۴	سوال چهارم.....
۵	سوال پنجم.....
۶	بخش پیاده سازی.....

بخش نوشتاری

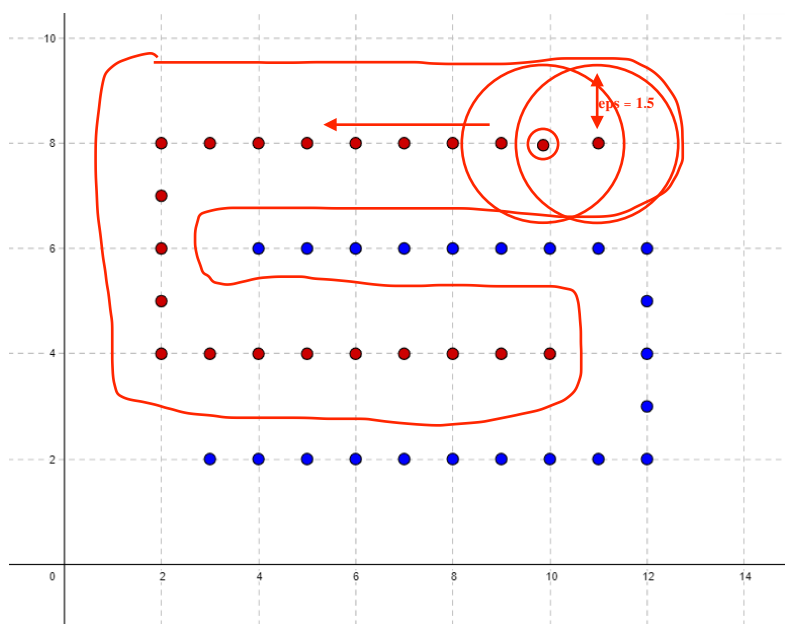
سوال اول

- الگوریتم kmeans را روی داده‌های شکل زیر با شرح مراحل اجرا کنید و خوشه‌ها را تعیین کنید.
- از نرم یک به عنوان معیار فاصله استفاده کنید.
 - تعداد خوشه‌ها را ۲ در نظر بگیرید و C_1 و C_2 مراکز ابتدایی‌اند.



سوال دوم

- یک الگوریتم مناسب برای خوشه‌بندی مجموعه داده‌ی زیر پیشنهاد دهید و توضیح دهید که به چه دلیل آن را انتخاب کردید. سپس با تعیین پارامترهای آن، عملکرد این الگوریتم را روی این داده‌ها تحلیل کنید.



سوال سوم

توضیح دهید که چگونه با ذخیره‌ی closed frequent itemsets می‌توان پشتیبانی مربوط به frequent itemset ها را تعیین کرد.

سوال چهارم

الگوریتم apriori را بر روی تراکنش‌های زیر اجرا کنید. تمامی مراحل تولید مجموعه آیت‌های کاندید را نشان دهید و در نهایت مجموعه آیت‌های پرتکرار را بدست آورید. همچنین تمامی قواعد انجمنی قابل تولید از مجموعه آیت‌ها را نوشته، آنهایی که مطمئن هستند را مشخص کرده و براساس میزان اطمینان مرتب کنید (آستانه پشتیبانی را ۳۳٪ و آستانه اطمینان را ۶۰٪ در نظر بگیرید).

شماره تراکنش	آیت‌ها
۱	سیب، پرتقال، موز
۲	انار، موز
۳	سیب، پرتقال، موز

۴	انار، پرتقال
۵	سیب، نارنگی
۶	سیب، نارنگی، انار

سوال پنجم

از ماتریس فاصله در جدول زیر برای انجام خوشه‌بندی سلسله مراتبی با لینک تک و کامل (min, max) استفاده کنید. نتایج خود را با کشیدن یک **دندروگرام** نشان دهید. در رسم باید به روشنی **ترتیب ادغام** نقاط نشان داده شود.

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

بخش پیاده‌سازی

در این بخش، به پیاده‌سازی الگوریتم‌هایی از مبحث خوشه‌بندی و قوانین انجمنی می‌پردازیم. به همراه تمرین دو فایل ipynb قرار داده شده است که شامل سوالات و نکات تمرین است. پرسش‌های این تمرین در قالب یک متن سبز و با پیش‌متن «پرسش» مطرح شده است. شما باید این سوالات را پاسخ دهید. دقت کنید که پاسخ‌های خود را نیز در قالب همان فایل ipynb بنویسید و از تغییر متغیرهای ثابت در تمرین اجتناب کنید.

این فایل‌ها در محیط گوگل کولب نوشته شده است. شما می‌توانید آن‌ها را به گوگل کولب خود انتقال دهید و فایل‌ها را ویرایش کنید. همچنین می‌توانید از محیط‌های دیگر، همچون `jupyter`، `vscode` و `notebook` و ... استفاده کنید. دقت کنید ممکن است در این محیط‌ها برخی از کتابخانه‌ها موجود نباشد که باید خودتان آن‌ها را نصب کنید.

برای تمرین کاهش حجم عکس، عکسی نیز به همراه تمرین قرار داده خواهد شد. برای کاهش حجم عکس، از آن عکس استفاده کنید.

برای تمرین قوانین انجمنی، فایل مجموعه داده به همراه تمرین قرار داده خواهد شد.

در انتها ما برای این بخش از شما دو فایل ipynb می‌خواهیم که به همراه باقی سوالات در قالب یک فایل zip ارسال می‌کنید.